AD-786 755

THE ECONOMICS OF DECISION MAKING

Thomas Richard Rice

Stanford University

Prepared for:

Office of Naval Research
Advanced Research Projects Agency
National Science Foundation

31 August 1974

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

AD786755

Research Report No. EES-DA-74-1
May 1, 1973 to August 31, 1974

THE ECONOMICS OF DECISION MAKING

THOMAS RICHARD RICE

DECISION ANALYSIS PROGRAM

Professor Ronald A. Howard
Principal Investigator

DEPARTMENT OF ENGINEERING-ECONOMIC SYSTEMS

Stanford University
Stanford, California 94305

SPONSORSHIPS

Advanced Research Projects Agency, Human Resources Research Office, ARPA Order No. 2449, monitored by Engineering Psychology Programs, Office of Naval Research, Contract No.N00014-67-A-0112-0077 (NR197-024)

National Science Foundation, NSF Grant GK-36491

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: EES-DA-74-1
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER: AD 786 755
4. TITLE (and Subtitle): "The Economics of Decision Making"
5. TYPE OF REPORT & PERIOD COVERED: Technical 5/1/73 to 8/31/74
6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Thomas Richard Rice
8. CONTRACT OR GRANT NUMBER(s): N00014-67-A-0112-0077
9. PERFORMING ORGANIZATION NAME AND ADDRESS: The Board of Trustees of the Leland Stanford Junior University, c/o Office of Research Administrator, Encina Hall, Stanford, California 94305
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 000000 ARPA Order #2449
11. CONTROLLING OFFICE NAME AND ADDRESS: Advanced Research Projects Agency, Human Resources Research Office, 1400 Wilson Blvd., Arlington, Virginia 22209
12. REPORT DATE: August 31, 1974
13. NUMBER OF PAGES: 111 pages
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Engineering Psychology Programs, Code 455, Office of Naval Research, 800 No. Quincy Street, Arlington, Virginia 22217
15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):
18. SUPPLEMENTARY NOTES: A dissertation submitted to the Department of Engineering-Economic Systems and the Committee on Graduate Studies of Stanford University in partial fulfillment of the Requirements for the Degree of Doctor of Philosophy, June 1974.
19. KEY WORDS (Continue on reverse side if necessary and identify by block number):

DECISION ANALYSIS	DECISION THEORY	APPROXIMATE ANALYSES
DECISION-MAKING	VALUE OF INFORMATION	QUADRATIC PROBLEM

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The design of a decision analysis is itself a complex decision problem. In theory, each aspect of analysis, encoding the probability density functions of state variables, encoding the von Neuman-Morgenstern utility function, and computing profit lotteries is an experiment. The results of the experiments, the data, are used to update the probabilities in the primary decision problem. The economic value of the experiment is the well known value of imperfect information. (continued on reverse)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The drawback to the theoretical approach is that the data are functions. Practical methods for encoding prior distributions over functions do not exist. Therefore, the traditional approach is to parameterize the data.

Our approach is unique because we show that for an interesting class of decision problems, arbitrary parameterization is not necessary. The value of any data depends probabilistically only on the prior covariances of the posterior means. For independent state variables this quantity reduces to an estimate of how much the mean of a probability density function will shift during an experiment.

With some limitations this result extends to the local risk aversion coefficient. The coefficient can be treated as if it were a state variable. The value of assessing the complete utility function is then proportional to the prior variance of the posterior coefficient. Once again encoding the potential mean shift is the key to the value of data generation.

The main result for computation is logically separate from the previous ones. The problem is to find the optimal quantization for a single decision variable. Monte Carlo samples from the profit or value function can be generated for any setting of the decision variable. For a fixed total sample size should we sample many times at a few decision points or a few times at many decision points? The answer is that fine quantization, implying many decision settings, is always superior. However, the expected loss from rough quantization is very small.

In the final chapter of the thesis we present flow charts which show how our results can be applied to the design of a practical decision analysis.

ACKNOWLEDGMENTS

My advisor, Professor Ronald A. Howard, has given me invaluable support and guidance throughout my graduate work at Stanford. Dr. James E. Matheson, Director of Stanford Research Institute's Decision Analysis Group, has acted as a second advisor. My interest in approximate analysis originated during an internship under Dr. Matheson, and my ideas have been focused through continuing discussions with him. Thanks are also due to Professor Donald A. Dunn for reviewing the manuscript and providing many helpful suggestions.

Discussions with Lee Merkhofer were particularly fruitful since his research interests overlap with mine. In fact, early discussions between the two of us resulted in the simultaneous and independent development of the expected value of clairvoyance for a quadratic value function. Theorem 1 of Chapter 2 is the extension of that basic result to imperfect information.

Bruce Judd, Lee Merkhofer and Steve Barrager were the nucleus of an unofficial review committee that donated many hours to improving the presentation.

My wife Ann's contribution is immeasurable. She has provided continuous editorial, emotional and financial support.

I am grateful to the National Science Foundation for the Fellowship which supported my early graduate work.

This research was partially supported by the Advanced Research Projects Agency of the Department of Defense, as monitored by the Office of Naval Research under Contract No. N00014-67-A-0112-0077. This research was also partially supported by the National Science Foundation under NSF Grant GK-36491.

ABSTRACT

The design of a decision analysis is itself a complex decision problem. In theory, each aspect of analysis, encoding the probability density functions of state variables, encoding the von Neuman-Morgenstern utility function, and computing profit lotteries is an experiment. The results of the experiments, the data, are used to update the probabilities in the primary decision problem. The economic value of the experiment is the well known value of imperfect information.

The drawback to the theoretical approach is that the data are functions. Practical methods for encoding prior distributions over functions do not exist. Therefore, the traditional approach is to parameterize the data.

Our approach is unique because we show that for an interesting class of decision problems, arbitrary parameterization is not necessary. The value of any data depends probabilistically only on the prior covariances of the posterior means. For independent state variables this quantity reduces to an estimate of how much the mean of a probability density function will shift during an experiment.

With some limitations this result extends to the local risk aversion coefficient. The coefficient can be treated as if it were a state variable. The value of assessing the complete utility function is then proportional to the prior variance of the posterior coefficient. Once again encoding the potential mean shift is the key to the value of data generation.

The main result for computation is logically separate from the

previous ones. The problem is to find the optimal quantization for a single decision variable. Monte Carlo samples from the profit or value function can be generated for any setting of the decision variable. For a fixed total sample size should we sample many times at a few decision points or a few times at many decision points? The answer is that fine quantization, implying many decision settings, is always superior. However, the expected loss from rough quantization is very small.

In the final chapter of the thesis we present flow charts which show how our results can be applied to the design of a practical decision analysis.

TABLE OF CONTENTS

page

LIST OF FIGURES

CHAPTER 1

OVERVIEW

1.0 Introduction

The design of a decision analysis is itself a complex decision problem. This dissertation addresses the analyst's decision of how much computation and assessment is economically justified for a given primary decision problem. The results are in two areas. In the first part of the thesis we extend decision theory to cover problems that can be approximated by Taylor series. These results apply to large decision problems where complete computation is infeasible. In the second part of the thesis we apply the results to the specific decisions of setting the levels of assessment and computation within a decision analysis.

The practical side of analysis, problem bounding and analytical design, has always been left to intuition. To handle extremely complex problems, a more formal approach is necessary. The analyst's skill at problem formulation will never be eliminated, but the approximate techniques developed in this dissertation should allow him to start with a very general representation of the problem and rationally eliminate the unimportant aspects.

1.1 Decision Analysis

Decision analysis is a practical discipline. It rests on twin foundations of decision theory and systems analysis. The reader of this dissertation will need at least an elementary knowledge of Bayesian decision theory. Excellent introductions to subjective probability and

risk preference are given in Howard [2]* and Raiffa [6]. Systems theory allows us to extend rational analysis to complex problems. The reader should understand the manipulation of matrices and optimization of functionals.

The state of information is a fundamental concept in decision analysis. The state of information that concerns us most frequently is $\mathcal{E}$, the decision maker's prior knowledge and experience. Another familiar state of information in decision analysis is clairvoyance $(C,\mathcal{E})$. The clairvoyant knows the exact value of any uncertain variable. In this dissertation we will normally be concerned with the augmented state of information $(D,\mathcal{E})$. If the data D contains no useful information, $(D,\mathcal{E})$ reduces to $\mathcal{E}$, and if the data is perfect information $(D,\mathcal{E})$ becomes $(C,\mathcal{E})$.

The relationship of the three states of information can be clarified using Howard's [2] decision analysis cycle. In Fig. 1.1 we associate $\mathcal{E}$ with the deterministic phase, $(D,\mathcal{E})$ with the probabilistic phase, and $(C,\mathcal{E})$ with the informational phase. Calling the initial phase deterministic is a mild misnomer since it is the basis for preliminary probabilistic estimates. The probability density function for a state variable can be approximated from the estimates of its mean and range. The profit lottery, the probability density function on the value, can be estimated from sensitivity data using Taylor series. The deterministic phase provides the given information for this paper. The probabilistic phase encompasses the encoding and computation that we wish to design. The informational phase is only of interest for its role in the three-part analogy.

*Numbers in square brackets refer to List of References.

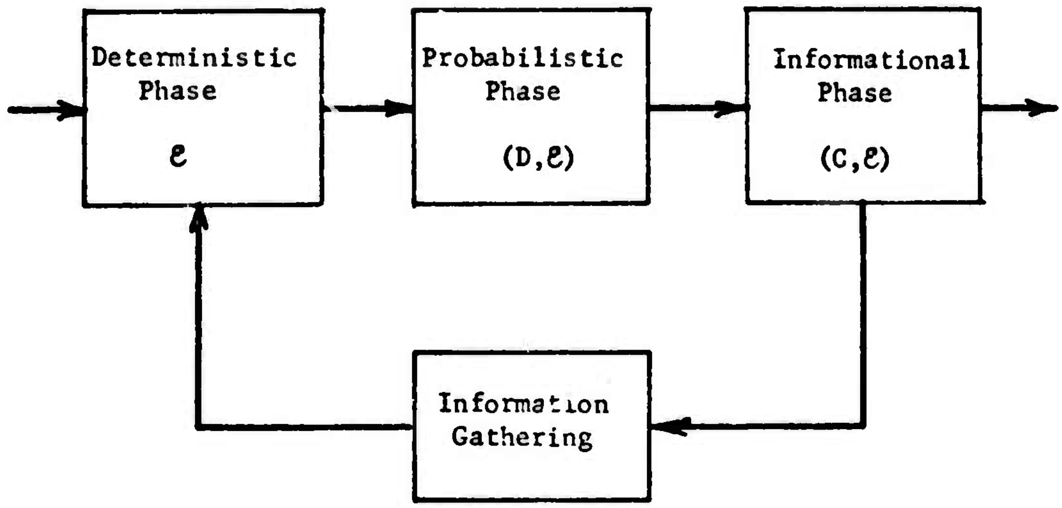


Figure 1.1 States of information within the decision analysis cycle

Related Work

We model computation and assessment as experiments. Almost all texts on decision theory present one or more special cases of the results in Chapters 2 and 3.

There are very few references that specifically address the problem of the design of a decision analysis. Howard [2] discusses the general philosophy. He points out that finding the right problem is as important as solving it. Raiffa and Schlaifer [7] introduce the use of parameters of probability distributions as random variables. Matheson [4] proposes a structure which is very similar to ours. Specifically, he introduces the concept that the purpose of analysis is to provide data to improve the state of information in the primary problem. The main difference between Matheson's work and ours is that we use an approximate value function for the primary problem. The approximation drastically reduces the required input, making it practical for application to complex decision problems.

Approximate value functions based on Taylor series are introduced in Howard [3]. Chapters 2 and 3 are an extension of Howard's structure.

1.2 Summary of Results

The thesis begins and ends with examples that illustrate the application of our theoretical results. The example at the start of Chapter 2 demonstrates that for certain problems determinisitic, rather than stochastic sensitivities are sufficient to calculate the value of clairvoyance. In Chapter 5 we return to the same example to illustrate the application of the results from Chapters 2, 3 and 4.

In Chapter 2 we solve the single-stage risk-indifferent decision

problem which has many decision variables related to many state variables through a second order value model. The results are exact for a quadratic value function and approximate for a complex value function that can be expanded in a Taylor series about the mean of the state variables and the deterministic optimum decision. The expression for the value of data can be decomposed into two parts. The first is a matrix representing the difference between closed and open loop sensitivities. Closed loop implies the ability to optimize the decision variables after the state variables are revealed. The second is the prior covariance of posterior means. For one state variable this quantity reduces to the prior variance of the posterior mean, a single parameter. This is a tremendous simplification over the general case in which the value of data depends on our prior estimate of the posterior probability distribution, a probability distribution over probability distributions.

In Chapter 3 we extend the results of Chapter 2 to include exponential risk aversion. The approximate value of clairvoyance derived in Section 3.2 is not useful for calculations because it involves third and fourth covariances. However, by considering special cases of the value of clairvoyance we derive criteria which must hold for the results of Chapter 2 to be valid. Finally, we calculate the loss from deliberate suppression of risk preference and the gain from introducing the decision maker's true utility function.

In Chapter 4 we address the question of discretizing a decision variable when the value lottery is generated by Monte Carlo simulation. The result is that a given number of random samples generates slightly

less expected error when the decision variable is finely discretized. Regardless of the discretization level, the expected error is approximately proportional to the total number of samples.

Chapter 5 is best summarized by Fig. 1.2. Each box represents a stage in the design of a decision analysis. Before we can apply our techniques, we need preliminary data. Then if the problem is suitable for approximate analysis, we consider the encoding, risk preference and computational decisions. For each state variable the encoding decision is whether to encode a complete probability density function or to use our preliminary estimate. The risk preference alternatives are to use a linear, exponential or general risk preference function. The computational alternatives are to stop after the preliminary analysis or to continue with a Monte Carlo simulation.

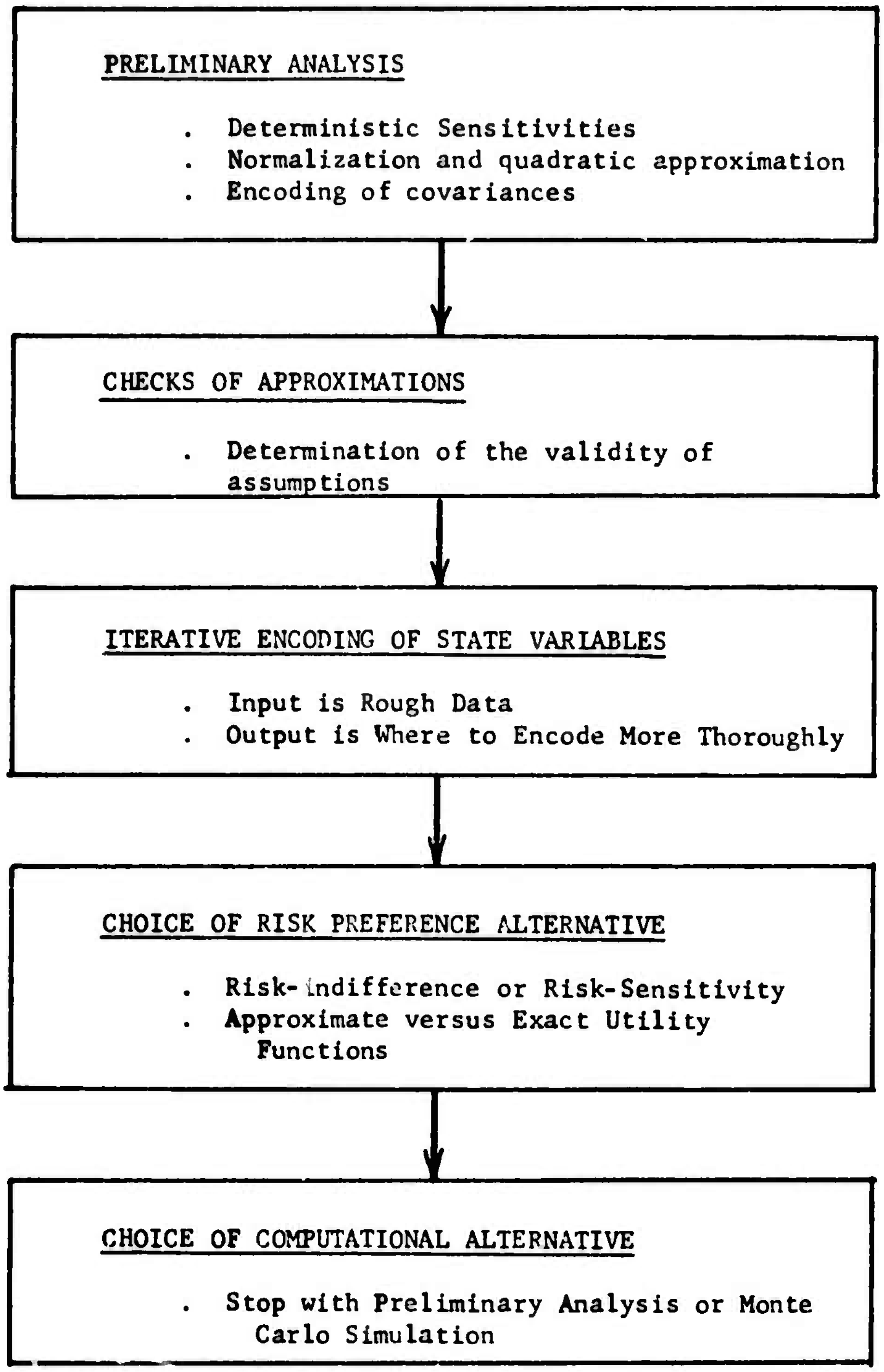


Figure 1.2 Summary of the economics of decision analysis

CHAPTER 2

VALUE OF ANALYSIS FOR THE RISK-INDIFFERENT DECISION MAKER

2.0 Introduction

This chapter begins with a simple example to introduce the concept of value of information. For the quadratic problem, which arises in practice when the value function can be approximated by a second-order Taylor series, we prove a general theorem for the value of data. This theorem is extended in Chapter 3 and applied in Chapter 5. At the end of this chapter we discuss how to handle non-quadratic problems and deliberate errors.

2.1 Preliminaries

In this section we introduce inferential notation and the general terminology required to describe a decision problem.

Notation

Inferential notation is well suited for this thesis because it explicitly conditions all probabilities on a state of information. The probability density function of a random variable x conditioned on the state of information $\mathcal{S}$ is denoted by

$$\{x|\mathcal{S}\} . \tag{2.1.1}$$

We use $\int_x$ as a generalized summation operator; thus the k^{th} moment of x is

$$\langle x^k|\mathcal{S}\rangle = \int_x x^k\{x|\mathcal{S}\} \tag{2.1.2}$$

whether x is continuous or discrete. Inferential notation can be nested. For example,

$$\{\langle x|\mathcal{S}_2\rangle|\mathcal{S}_1\} \tag{2.1.3}$$

implies that the mean of $\{x|\mathcal{S}_2\}$ is a random variable given only $\mathcal{S}_1$.

In addition to inferential notation, we use the following matrix symbols:

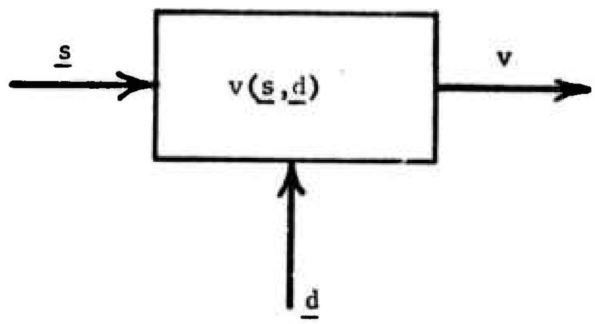
$\underline{a}$ or $a_i]$	The underscored lower case letter denotes a column vector with element a_i.
$\underline{A}$ or $[a_{ij}]$	The underscored capital letter denotes a square matrix with element a_{ij}.
$\underline{a}'$ or $\underline{A}'$	The prime denotes transposition.
$\langle\underline{a}\|\mathcal{S}\rangle$ or $\langle a_i\|\mathcal{S}\rangle]$	A probabilistic operation is applied to each component of a vector.

The Basic Decision Problem

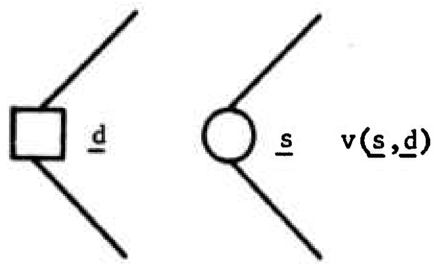
The deterministic model illustrated in Fig. 2.1 relates the three elements of the basic decision problem. The decision variables $\underline{d}$ are set by the decision maker. The state variables $\underline{s}$ are set by nature. The value v is the output measure that we want to maximize. If both $\underline{s}$ and $\underline{d}$ are known, we denote the decision that maximizes the value function $\hat{\underline{d}}(\underline{s})$:

$$\hat{\underline{d}}(\underline{s}) = \max_{\underline{d}}{}^{-1} v(\underline{s},\underline{d}) \tag{2.1.4}$$

However, in the basic decision problem illustrated in Fig. 2.1b, $\underline{d}$ must set before $\underline{s}$ is observed. The possible outcomes are described by the probability density function $\{\underline{s}|\mathcal{E}\}$, where $\mathcal{E}$ is the state of information that represents the decision maker's prior knowledge and experience.



(a) The deterministic model



(b) The probability tree

Figure 2.1 Description of the Basic Decision Problem

We assume that $\underline{s}$ is independent of $\underline{d}$ in the sense that

$$\{\underline{s}|\underline{d},\mathcal{E}\} = \{\underline{s}|\mathcal{E}\} . \tag{2.1.5}$$

This assumption is not restrictive. When the state variables are dependent on the decision variables the problem can normally be reformulated so that the dependence appears in the value function. The example in the following section illustrates how state variables can be made probabilistically independent of decision variables.

The basic decision problem under uncertainty is to maximize the expectation of v :

$$\max_{\underline{d}} \int_{\underline{s}} v(\underline{s},\underline{d}) \{\underline{s}|\mathcal{E}\} \tag{2.1.6}$$

The expansion rule from elementary probability theory is

$$\langle x|\mathcal{E}\rangle = \int_{y} \langle x|y,\mathcal{E}\rangle \{y|\mathcal{E}\} . \tag{2.1.7}$$

Using this rule, we can show that the inferential symbol for the expectation in (2.1.6) is $\langle v|\underline{d},\mathcal{E}\rangle$:

$$\langle v|\underline{d},\mathcal{E}\rangle = \int_{\underline{s}} \langle v|\underline{s},\underline{d},\mathcal{E}\rangle \{\underline{s}|\mathcal{E}\} \tag{2.1.8}$$

The expectations in (2.1.6) and (2.1.8) are the same since the expected value of v given $\underline{s}$ and $\underline{d}$ is deterministically $v(\underline{s},\underline{d})$.

We define $\hat{\underline{d}}(\mathcal{E})$ as the decision vector that maximizes the expected value of v :

$$\hat{\underline{d}}(\mathcal{E}) = \max_{\underline{d}}^{-1} \langle v|\underline{d},\mathcal{E}\rangle \tag{2.1.9}$$

If $\mathcal{S}$ represents some possible future state of information, we define $\underline{d}^*(\mathcal{S})$ as the intent to use $\hat{\underline{d}}(\mathcal{S})$ when $\mathcal{S}$ becomes available.

The Value of Information

Suppose that an analysis or experiment will provide some data D . Then $(D,\mathcal{E})$ represents an improved state of information. We define the expected value of the data $\langle v_D|\mathcal{E}\rangle$:

$$\langle v_D|\mathcal{E}\rangle = \langle v|\underline{d}^*(D,\mathcal{E}),\mathcal{E}\rangle - \langle v|\underline{d}^*(\mathcal{E}),\mathcal{E}\rangle \tag{2.1.10}$$

Since $\mathcal{E}$ is our prior information, $\hat{\underline{d}}(\mathcal{E})$ is known and thus

$$\langle v|\underline{d}^*(\mathcal{E}),\mathcal{E}\rangle = \langle v|\hat{\underline{d}}(\mathcal{E}),\mathcal{E}\rangle \quad . \tag{2.1.11}$$

The first term in (2.1.10) is the key to the value of data. Given the data D we would find

$$\hat{\underline{d}}(D,\mathcal{E}) = \max_{\underline{d}}{}^{-1}\langle v|\underline{d},D,\mathcal{E}\rangle \ , \tag{2.1.12}$$

which would result in the posterior expected value $\langle v|\hat{\underline{d}}(D,\mathcal{E}),D,\mathcal{E}\rangle$. However, before D is revealed we must compute the prior expectation of this quantity:

$$\langle v|\underline{d}^*(D,\mathcal{E}),D,\mathcal{E}\rangle = \langle\langle v|\hat{\underline{d}}(D,\mathcal{E}),D,\mathcal{E}\rangle|\mathcal{E}\rangle \tag{2.1.13}$$

2.2 The Entrepreneur's Problem, an Example

The expected value of data is a very useful concept in applied decision analysis. The example of this section demonstrates its importance. In the remainder of the chapter we examine conditions under which the example can be generalized to more complex problems.

The example, The Entrepreneur's Problem, was originally formulated by Howard [3]. Our methodology differs from Howard's, but our numerical

results are the same. The reader need not be familiar with [3] to understand the example.

Description of the Model

The Entrepreneur's Problem is illustrated by the schematic tree of Fig. 2.2a. The entrepreneur must decide at what price p to sell his new product. His profit π is the revenue, price p times quantity sold q, minus the cost c. Deterministically, the quantity sold is related to price through the demand curve $q(p)$. The total cost is related to the quantity sold and consequently the price through the cost function $c(q(p))$.

The problem is simplified by assuming that given the prior state of information $\mathcal{E}$, c and q are probabilistically independent of p and of each other. The quantity Δq is defined as the difference between the actual demand and the nominal demand $q(p)$. Likewise, Δc is the difference between actual and nominal cost. The independence assumption implies that

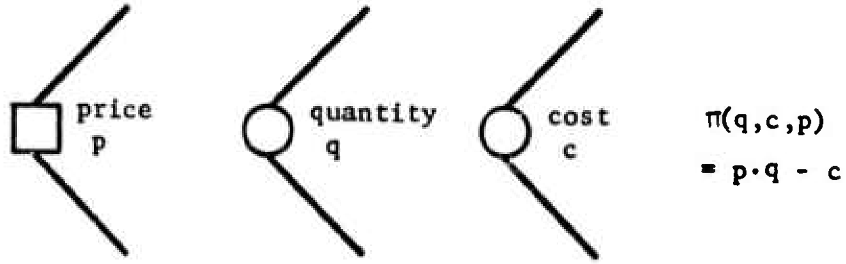
$$\{\Delta q, \Delta c \,|\, p, \mathcal{E}\} = \{\Delta q \,|\, \mathcal{E}\}\{\Delta c \,|\, \mathcal{E}\} . \tag{2.2.1}$$

Both Δc and Δq are assumed to have zero mean:

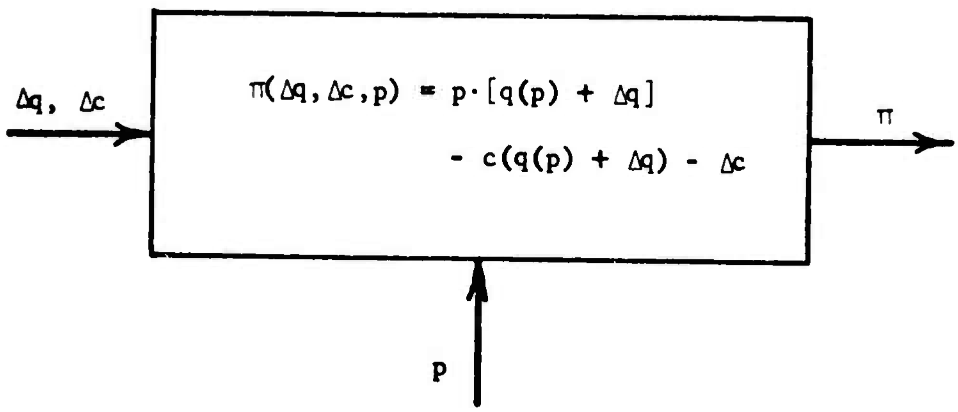
$$\langle \Delta q \,|\, \mathcal{E} \rangle = 0 \tag{2.2.2}$$

$$\langle \Delta c \,|\, \mathcal{E} \rangle = 0 \tag{2.2.3}$$

Using these simplifications we can modify the deterministic model as shown in Fig. 2.2.b. The demand and cost functions are incorporated into the model, leaving Δq, Δc and p as the input variables. The modified model is an example of the basic decision problem from Section 2.1. The presentation can be simplified without loss of generality by using the reduced random variables



(a) The probabilistic model



(b) The simplified deterministic model

Figure 2.2 The Entrepreneur's Problem

$$q_r = \Delta q / \overset{v}{<} q|\mathcal{E}> \tag{2.2.4}$$

$$c_r = \Delta c / \overset{v}{<} c|\mathcal{E}> . \tag{2.2.5}$$

Deterministic Data

The first step in analyzing the Entrepreneur's Problem is to perform deterministic sensitivities. The inputs to the deterministic model of Fig. 2.2b are varied, and the resulting change in profit is observed. The sensitivity plots, Figs. 2.3 through 2.6 serve as a numerical description of the problem.

The first sensitivity is to price. In Fig. 2.3 we see that the deterministic optimum price $\hat{p}_o$ is 24.1 . As price is raised or lowered by 10, the profit drops from 198 to 14. The three points are sufficient to determine a quadratic approximation to the price sensitivity $\pi(p)$:

$$\pi(p) = \pi(q_r, c_r, p) \tag{2.2.7}$$

where

$$q_r = c_r = 0 \tag{2.2.8}$$

More compactly we express this sensitivity as

$$\pi(p) = \pi(0,0,p) . \tag{2.2.9}$$

The sensitivities to quantity are shown in Fig. 2.4. Once again we use three points to find a quadratic approximation. The open loop sensitivity is performed by holding c_r and p constant while q_r varies:

$$\pi_o(q_r) = \pi(q_r, 0, \hat{p}_o) \tag{2.2.10}$$

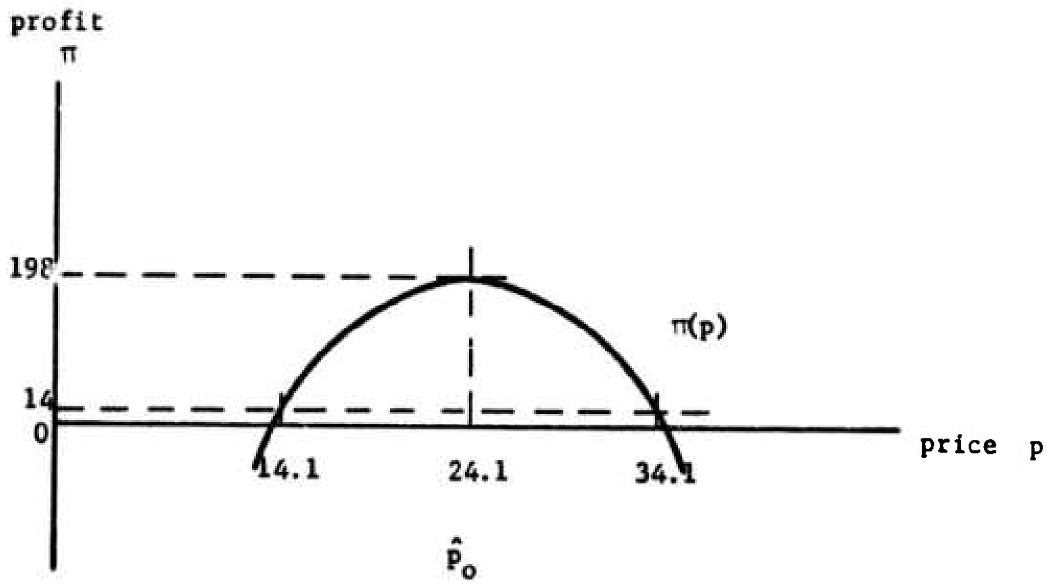


Figure 2.3 Deterministic Sensitivity to price

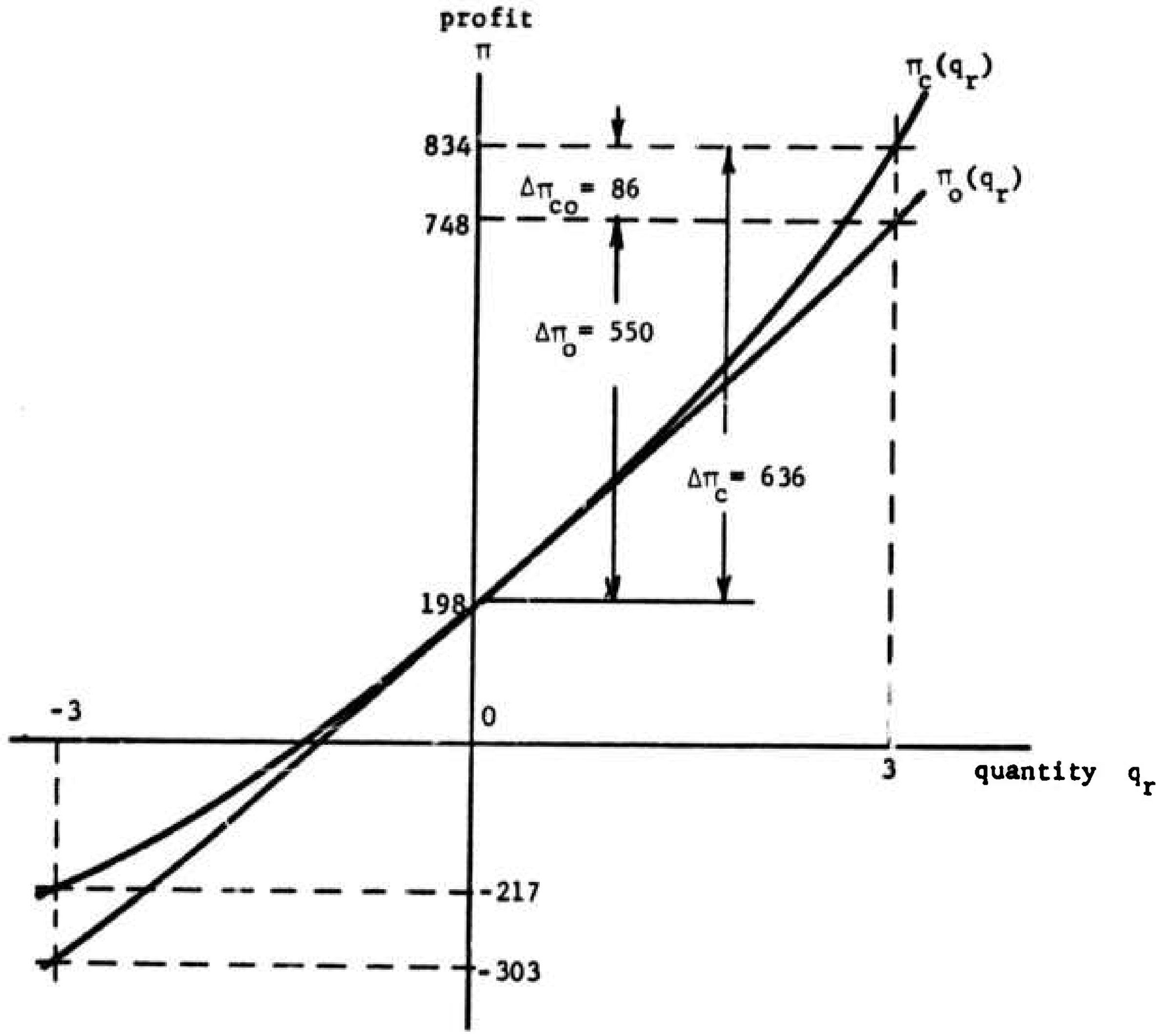


Figure 2.4 Open and closed loop sensitivities for quantity

To understand the generation of the closed loop sensitivity we consider the point where the quantity is 3 standard deviations above the mean. The open loop profit at this point is 748, an increase of 555 over the profit at $q_r = 0$. We denote the open loop increase $\Delta\pi_o$.

Turning to Fig. 2.5, we see that $\Delta\pi_o$ corresponds to the increase in profit with the decision fixed at $\hat{p}_o$:

$$\Delta\pi_o = \pi(3,0,\hat{p}_o) - \pi(0,0,\hat{p}_o) \qquad (2.2.11)$$

We see that 748 is just one point on the top curve in Fig. 2.5, $\pi(3,0,p)$, the price sensitivity for $q_r = 3$. The maximum of this curve is

$$834 = \max_p \pi(3,0,p) . \qquad (2.2.12)$$

Returning to Fig. 2.4, we see that 834 is also the value of the closed loop sensitivity to quantity evaluated at $q_r = 3$. Therefore, the closed loop sensitivity to quantity is the change in profit given the opportunity to reoptimize profit after the quantity is revealed. We see from either Fig. 2.4 or Fig. 2.5 that the closed loop change $\Delta\pi_c$ can be decomposed into the open loop change $\Delta\pi_o$ plus the compensation $\Delta\pi_{co}$. The latter term is due to the change in decision $\Delta\hat{p}_{co}$.

The unique characteristic of the cost sensitivities of Fig. 2.6 is that the open and closed loop curves coincide. This indicates that the decision is insensitive to cost. In effect the entrepreneur has written a blank check to his creditors. He is uncertain about the differential cost Δc, but he cannot influence its resolution.

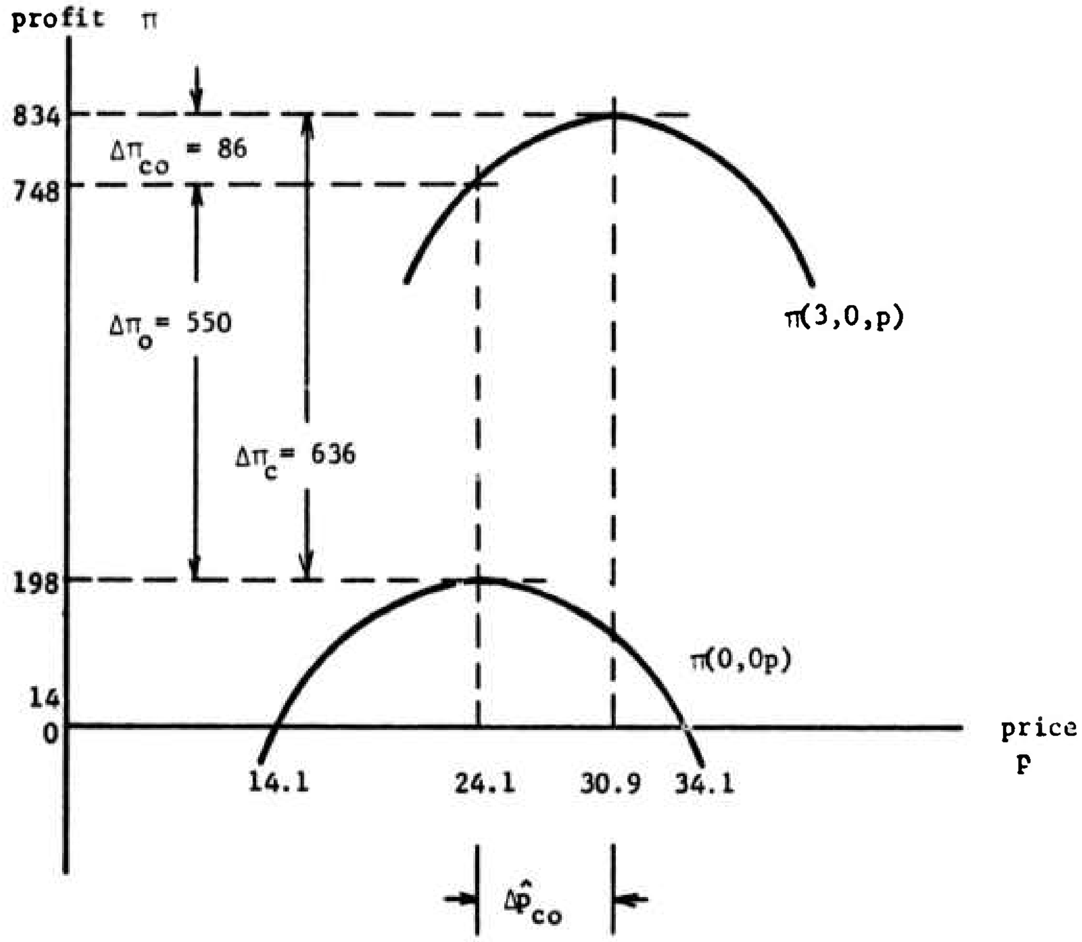


Figure 2.5 The generation of the closed loop sensitivity to quantity

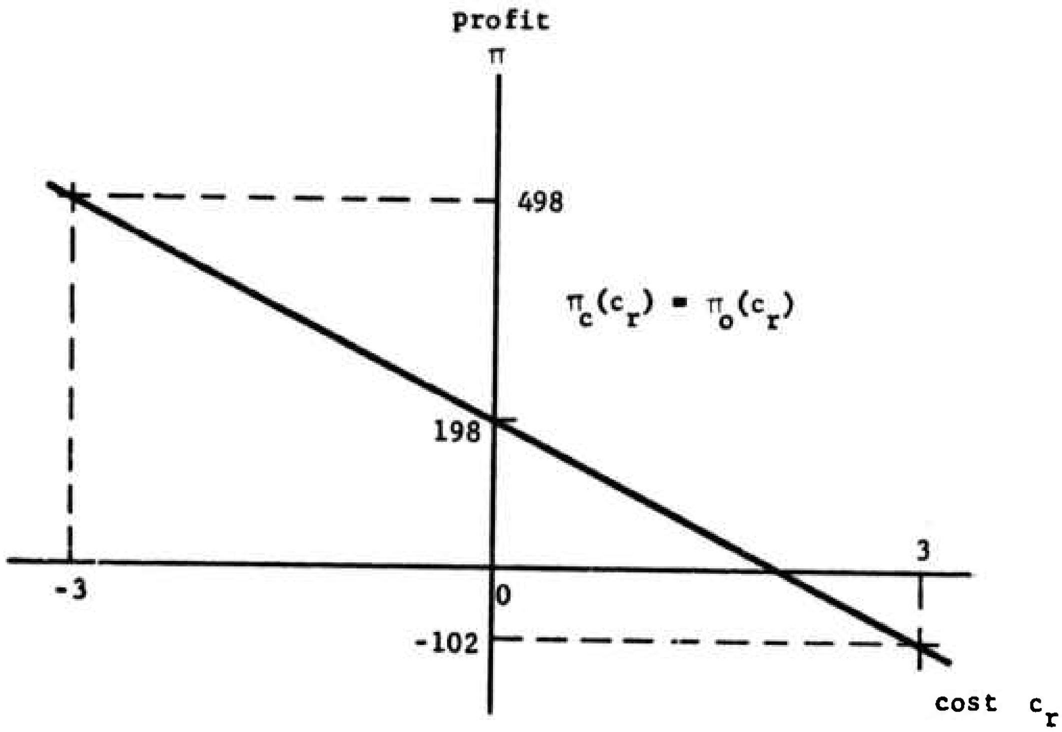


Figure 2.6 Open and closed loop sensitivities for cost

Value of Information for the Entrepreneur's Problem

We now use the problem formulation and the sensitivity data to solve the Entrepreneur's Problem and to calculate the value of clairvoyance on the state variables. As we shall see in the following section these calculations are exact for a quadratic problem. The value of clairvoyance is interesting because it is prototypal of the approximate value of data for more complex problems.

Finding the prior optimum decision $\hat{p}(\mathcal{E})$ is straightforward. The entrepreneur's profit function is quadratic in the state and decision variables. As we shall see in the next section this implies that the deterministic and probabilistic optimum decision coincide:

$$\hat{p}(\mathcal{E}) = \hat{p}_o \tag{2.2.13}$$

To find the value of clairvoyance on the state variables, we specialize (2.1.10) to

$$\langle v_C|\mathcal{E}\rangle = \langle \pi|p^*(C,\mathcal{E}),\mathcal{E}\rangle - \langle \pi|\hat{p}(\mathcal{E}),\mathcal{E}\rangle \ . \tag{2.2.14}$$

Since clairvoyance $(C,\mathcal{E})$ is equivalent to exact knowledge of q_r and c_r we can expand (2.2.14) to

$$\langle v_C|\mathcal{E}\rangle = \langle\langle \pi|\hat{p}(q_r,c_r),q_r,c_r,\mathcal{E}\rangle - \langle \pi|\hat{p}(\mathcal{E}),q_r,c_r,\mathcal{E}\rangle|\mathcal{E}\rangle \ . \tag{2.2.15}$$

The inner expression of (2.2.15) can be evaluated from sensitivity data. Since changes in cost do not affect price, the first term reduces to

$$\langle \pi|\hat{p}(q_r),q_r,c_r,\mathcal{E}\rangle = \pi(q_r,c_r,p(q_r)) = \pi(q_r,0,p(q_r)) - \Delta c \ . \tag{2.2.16}$$

The decomposition of the profit is possible because the cost is additive. Likewise, the second term is

$$<\pi|\hat{p}(\mathcal{E}),q_r,c_r,\mathcal{E}> = \pi(q_r,c_r,\hat{p}_o) = \pi(q_r,0,\hat{p}_o) - \Delta c \ . \qquad (2.2.17)$$

Subtracting (2.2.17) from (2.2.16) we find that the inner term is the compensation $\Delta\pi_{co}(q_r)$, the difference between $\pi_c(q_r)$ and $\pi_o(q_r)$ in Fig. 2.4. Therefore, the value of clairvoyance is the expected value of compensation:

$$<v_c|\mathcal{E}> = \ <\pi(q_r,0,p(q_r)) - \pi(q_r,0,\hat{p}_o)|\mathcal{E}> = <\Delta\pi_{co}(q_r)|\mathcal{E}> \qquad (2.2.18)$$

Figure 2.7 illustrates the expected value of compensation:

$$<\pi_{co}(q_r)|\mathcal{E}> = \int_{q_r} \{q_r|\mathcal{E}\} \ \Delta\pi_{co}(q_r) \qquad (2.2.19)$$

Using the data from Fig. 2.4 the compensation is

$$\Delta\pi_{co}(q_r) = 9.5 \ q_r^2 \ . \qquad (2.2.20)$$

Substituting (2.2.20) into (2.2.19) and recalling that the reduced variable has zero mean and unit variance, the value of clairvoyance is

$$<v_c|\mathcal{E}> = 9.5 \int_{q_r} \{q_r|\mathcal{E}\} \ q_r^2 = 9.5 \ . \qquad (2.2.21)$$

The value of clairvoyance is the curvature of the closed loop sensitivity to the reduced state variable at $q_r = 0$. In the next section we show that the curvature of the closed loop sensitivity is important regardless of the source of the data.

2.3 The Value of Data for a Decision Problem with a Quadratic Value Function

In this section we derive the exact value of data for a quadratic value function. After we state the result and prove it, we suggest how

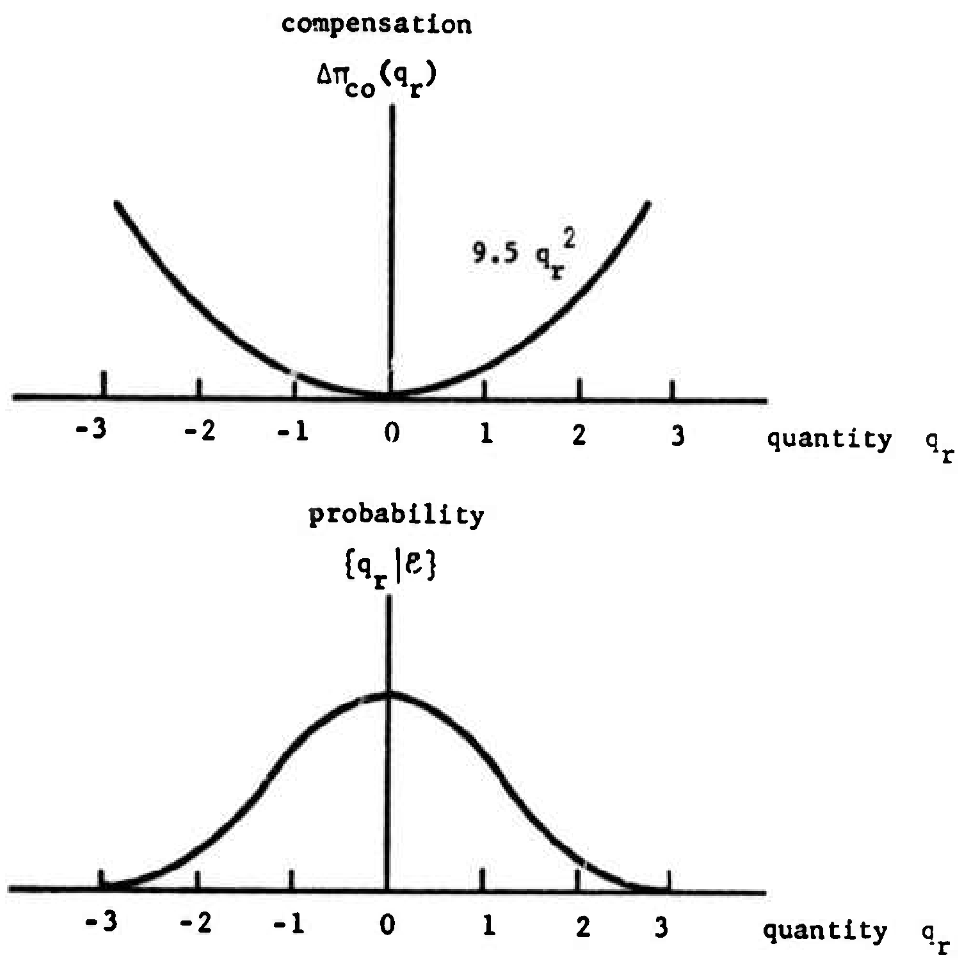


Figure 2.7 The two components of the value of clairvoyance

to apply it to non-quadratic problems. The theorem of this section is extended in Chapter 3 and applied in Chapter 5.

The deterministic model is illustrated in Fig. 2.8. The value function $v(\underline{s},\underline{d})$ is quadratic in the state vector $\underline{s}$ and the decision vector $\underline{d}$. We will normalize the state variables to have zero mean, and the decision variables to be zero at the deterministic maximum:

$$<\underline{s}|\mathcal{E}> = \underline{0} \tag{2.3.1}$$

$$\hat{\underline{d}}_o = \max_{\underline{d}}{}^{-1} v(<\underline{s}|\mathcal{E}>,d) = \underline{0} \tag{2.3.2}$$

These assumptions reduce algebraic complexity without sacrificing generality.

We write the quadratic value function as

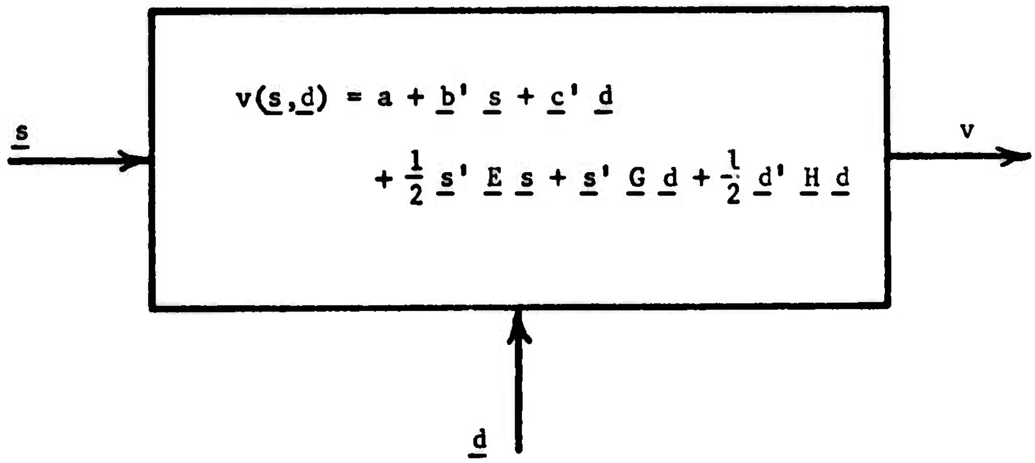
$$v(\underline{s},\underline{d}) = a + \underline{b}'\underline{s} + \underline{c}'\underline{d} + \frac{1}{2}\,\underline{s}'\underline{E}\,\underline{s} + \underline{s}'\underline{G}\,\underline{d} + \frac{1}{2}\,\underline{d}'\underline{H}\,\underline{d}\ . \tag{2.3.3}$$

The second-order necessary and sufficient conditions for $v(\underline{s},\underline{d})$ to have a maximum at $<\underline{s}|\mathcal{E}>$ and $\hat{\underline{d}}_o$ are that the gradient of v with respect to $\underline{d}$ $\nabla v(<\underline{s}|\mathcal{E}>,\hat{\underline{d}}_o)$ be zero and that the Hessian of v with respect to $\underline{d}\,\nabla^2 v<\underline{s}|\mathcal{E}>,\underline{d})$ be negative definite. Using (2.3.1) and (2.3.2) the gradient and Hessian at $<\underline{s}|\mathcal{E}>$ and $\hat{\underline{d}}_o$ are defined as

$$\nabla v(<\underline{s}|\mathcal{E}>,\hat{\underline{d}}_o) = \left\lfloor \frac{\partial v(\underline{0},\underline{0})}{\partial d_i} \right\rfloor \tag{2.3.4}$$

$$\nabla^2 v(<\underline{s}|\mathcal{E}>,\hat{\underline{d}}_o) = \left[\frac{\partial^2 v(\underline{0},\underline{0})}{\partial d_i\,\partial d_j}\right]. \tag{2.3.5}$$

Applying (2.3.4) and (2.3.5) to the definition of $v(\underline{s},\underline{d})$ (2.3.3), we have:



Notation

$'$	Transpose of a matrix
a	Constant
$\underline{b},\underline{c}$	Constant vectors
$\underline{s}$	State variable vector
$\underline{d}$	Decision variable vector
$\underline{E},\underline{G},\underline{H}$	Constant square matrices

Figure 2.3 The quadratic value model

$$\nabla v(\underline{0},\underline{0}) = \underline{c}' + \underline{0}'\underline{H} \tag{2.3.6}$$

$$\nabla^2 v(\underline{0},\underline{0}) = \underline{H} \tag{2.3.7}$$

Since the gradient in (2.3.6) must be the zero vector, our assumptions imply that $\underline{c}$ must also be the zero vector. From (2.3.7) we see that the Hessian does not vary with $\underline{s}$ and $\underline{d}$ for the quadratic. Therefore if the deterministic optimum $\hat{\underline{d}}_o$ exists, $\underline{H}$ is negative definite and the value function has a global maximum with respect to $\underline{d}$ for any state vector $\underline{s}$.

Chronologically, we receive the data about the state variables. Then we set the decision vector, and finally nature sets the state variables. The state variables are independent of the decision variables but not necessarily independent of each other. We assume that the decision maker is risk-indifferent so that maximizing the value function is equivalent to maximizing the decision maker's von Neumann-Morgenstern utility function.

With these preliminaries we can state the theorem:

<u>THEOREM</u>: For the quadratic value function

$$v(\underline{s},\underline{d}) = a + \underline{b}'\underline{s} + \frac{1}{2}\,\underline{s}'\underline{E}\,\underline{s} + \underline{s}'\underline{G}\,\underline{d} + \frac{1}{2}\,\underline{d}'\underline{H}\,\underline{d}\ , \tag{2.3.8}$$

where the Hessian $\underline{H}$ is negative definite, the value of any data D is

$$\langle v_D|\mathcal{E}\rangle = -\frac{1}{2}\langle\langle\underline{s}|D,\mathcal{E}\rangle'\ \underline{G}\ \underline{H}^{-1}\underline{G}'\ \langle\underline{s}|D,\mathcal{E}\rangle|\mathcal{E}\rangle\ . \tag{2.3.9}$$

<u>PROOF</u>: From (2.1.10) the value of the data is

$$\langle v_D|\mathcal{E}\rangle = \langle v|\underline{d}^*(D,\mathcal{E}),D,\mathcal{E}\rangle - \langle v|\underline{d}^*(\mathcal{E}),\mathcal{E}\rangle\ . \tag{2.3.10}$$

The proof is in two parts corresponding to the two terms of (2.3.10). First we determine the prior maximum expected value $\langle v|\underline{d}^*(\mathcal{E}),\mathcal{E}\rangle$; then we determine the expected value given the opportunity to maximize after the data is received $\langle v|\underline{d}^*(D,\mathcal{E}),D,\mathcal{E}\rangle$.

To find $\langle v|\underline{d}^*(\mathcal{E}),\mathcal{E}\rangle$ we start with the prior expected value $\langle v|\underline{d},\mathcal{E}\rangle$. Recalling that the expected values of the state variables are all zero, the prior expectation of (2.3.8) is

$$\langle v|\underline{d},\mathcal{E}\rangle = a + \frac{1}{2}\langle \underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle + \frac{1}{2}\underline{d}'\underline{H}\,\underline{d} \; . \qquad (2.3.11)$$

The first-order necessary condition for $\langle v|\hat{\underline{d}}(\mathcal{E}),\mathcal{E}\rangle$ to be an unconstrained maximum is that the gradient be zero at $\hat{\underline{d}}(\mathcal{E})$:

$$\nabla \langle v|\hat{\underline{d}}(\mathcal{E}),\mathcal{E}\rangle = \underline{0}' \qquad (2.3.12)$$

Taking the gradient of (2.3.11) and setting it to zero, we have

$$\hat{\underline{d}}'(\mathcal{E})\underline{H} = \underline{0}' \; . \qquad (2.3.13)$$

Since $\underline{H}$ is negative definite, $\hat{\underline{d}}(\mathcal{E})$ must be the zero vector. Therefore (2.3.11) becomes

$$\langle v|\hat{\underline{d}}(\mathcal{E}),\mathcal{E}\rangle = a + \frac{1}{2}\langle \underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle \; . \qquad (2.3.14)$$

Returning to the first term in (2.3.10), the expected value given data D is

$$\langle v|\underline{d},D,\mathcal{E}\rangle = a + \underline{b}'\langle \underline{s}|D,\mathcal{E}\rangle + \frac{1}{2}\langle \underline{s}'\underline{E}\,\underline{s}|D,\mathcal{E}\rangle + \langle \underline{s}|D,\mathcal{E}\rangle'\underline{G}\,\underline{d}$$
$$+ \frac{1}{2}\underline{d}'\underline{H}\,\underline{d} \; . \qquad (2.3.15)$$

Maximizing (2.3.15) with respect to $\underline{d}$ we have

$$\nabla <v|\hat{\underline{d}}(D,\mathcal{E}),D,\mathcal{E}> = <\underline{s}|D,\mathcal{E}>'\underline{G} + \underline{d}'\underline{H} = \underline{0}' \ . \tag{2.3.16}$$

Equation (2.3.16) implies that

$$\hat{\underline{d}}(D,\mathcal{E}) = - \underline{H}^{-1}\underline{G}' <\underline{s}|D,\mathcal{E}> \ . \tag{2.3.17}$$

Substituting (2.3.17) into (2.3.15), we have

$$\begin{aligned} <v|\hat{\underline{d}}(D,\mathcal{E}),D,\mathcal{E}> &= a + \underline{b}'<\underline{s}|D,\mathcal{E}> + \frac{1}{2} <\underline{s}'\underline{E}\ \underline{s}|D,\mathcal{E}> \\ &\quad - <\underline{s}|D,\mathcal{E}>'\underline{G}\ \underline{H}^{-1}\underline{G}' <\underline{s}|D,\mathcal{E}> + \frac{1}{2} <\underline{s}|D,\mathcal{E}>'\underline{G}\ \underline{H}^{-1}\underline{G} <\underline{s}|D,\mathcal{E}> \\ &= a + \underline{b}'<\underline{s}|D,\mathcal{E}> + \frac{1}{2} <\underline{s}'\underline{E}\ \underline{s}|D,\mathcal{E}> \\ &\quad - \frac{1}{2} <\underline{s}|D,\mathcal{E}>'\underline{G}\ \underline{H}^{-1}\underline{G}' <\underline{s}|D,\mathcal{E}> \ . \end{aligned} \tag{2.3.18}$$

Recalling (2.1.13), the next step is to take the prior expectation of (2.3.18). We shall consider each term separately. Of course, expectation does not affect the value of the constant a. The prior expectation of the posterior mean is the prior mean :

$$<<\underline{s}|D,\mathcal{E}>|\mathcal{E}> = <\underline{s}|\mathcal{E}> \tag{2.3.19}$$

Equation (2.3.19) is a direct application of the definition of conditional probability. Likewise, the third term becomes

$$<<\underline{s}'\underline{E}\ \underline{s}|D,\mathcal{E}>|\mathcal{E}> = <\underline{s}'\underline{E}\ \underline{s}|\mathcal{E}> \ . \tag{2.3.20}$$

Applying these results to (2.3.18), we have

$$<v|\underline{d}^*(D,\mathcal{E}),\mathcal{E}> = a + \frac{1}{2} <\underline{s}'\underline{E}\ \underline{s}|\mathcal{E}> - \frac{1}{2}<<\underline{s}|D,\mathcal{E}>'\underline{G}\ \underline{H}^{-1}\underline{G}' <\underline{s}|D,\mathcal{E}>|\mathcal{E}> \ . \tag{2.3.21}$$

Finally, subtracting (2.3.14) from (2.3.21) the result is

$$\langle v_D|\mathcal{E}\rangle = -\frac{1}{2}\langle\langle \underline{s}|D,\mathcal{E}\rangle' \underline{G}\ \underline{H}^{-1}\underline{G}'\ \langle \underline{s}|D,\mathcal{E}\rangle|\mathcal{E}\rangle . \quad \text{Q.E.D.} \qquad (2.3.22)$$

Special Cases of the Theorem That Appear in the Literature

Three special cases of the theorem (2.3.9) appear in the literature. Howard [3, p. 518] treats the case where $\underline{H}$ is diagonal and the data D is clairvoyance. DeGroot [1, p. 234] solves for $\underline{d}$, the estimate of the random variable $\underline{s}$ which minimizes a quadratic loss function. In our notation his problem is the case where

$$\underline{E} = -\underline{H} ; \qquad (2.3.23)$$

a, $\underline{b}$, and $\underline{G}$ are zero; and D is clairvoyance. Raiffa and Schlaifer [7, p. 188] present the one-dimensional estimation problem without requiring the data to be clairvoyance.

2.4 Discussion of the Value of Data for the Quadratic Problem

An alternate expression for the theorem (2.3.9) is:

$$\langle v_D|\mathcal{E}\rangle = -\frac{1}{2} \text{ trace } \underline{E}_{co}\underline{C}_D \qquad (2.4.1)$$

where

$$\underline{E}_{co} = \underline{G}\ \underline{H}^{-1}\underline{G}' \qquad (2.4.2)$$

$$\underline{C}_D = \left[\langle\langle s_i|D,\mathcal{E}\rangle\ \langle s_j|D,\mathcal{E}\rangle|\mathcal{E}\rangle\right] \qquad (2.4.3)$$

The trace of a matrix is the sum of its diagonal elements. The value of data has two major components. The basic decision problem is specified by $\underline{E}_{co}$, and the experiment is described by $\underline{C}_D$.

We consider $\underline{E}_{co}$ and $\underline{C}_D$ briefly for the general case. Then for

the case that is most common we discuss how $\underline{E}_{co}$ and $\underline{C}_D$ could be generated. Finally, we return to the Entrepreneur's Problem to illustrate the encoding of $\underline{C}_D$.

$\underline{E}_{co}$ follows directly from (2.4.2) for a true quadratic value function since the matrices $\underline{G}$ and $\underline{H}$ are specified. For a problem that is approximately quadratic, $\underline{G}$ and $\underline{H}$ can be found by expanding $v(\underline{s},\underline{d})$ in a Taylor series about the point $(\langle\underline{s}|\mathcal{E}\rangle, \hat{\underline{d}}(\mathcal{E})) = (\underline{0},\underline{0})$:

$$v(\underline{s},\underline{d}) = v(\underline{0},\underline{0}) + \underbrace{\frac{\partial v}{\partial s_i}}\,\underline{s} + \frac{1}{2}\,\underline{s}'\left[\frac{\partial^2 v}{\partial s_i\,\partial s_j}\right]\underline{s}$$

$$+ \underline{s}'\left[\frac{\partial^2 v}{\partial s_i\,\partial d_j}\right]\underline{d} + \frac{1}{2}\,\underline{d}'\left[\frac{\partial^2 v}{\partial d_i\,\partial d_j}\right]\underline{d}\ . \qquad (2.4.4)$$

The partial derivatives are all evaluated at the point $(\underline{0},\underline{0})$. Comparing (2.4.4) with (2.3.8), we see that $\underline{G}$ and $\underline{H}$ must be matrices of partial derivatives:

$$\underline{G} = \left[\frac{\partial^2 v}{\partial s_i\,\partial d_j}\right] \qquad (2.4.5)$$

$$\underline{H} = \left[\frac{\partial^2 v}{\partial d_i\,\partial d_j}\right] \qquad (2.4.6)$$

The partial derivatives at the operating point $(\underline{0},\underline{0})$ can be approximated from open loop sensitivities. One joint sensitivity is required for each possible pair of state and decision variables and for each possible pair of decision variables.

The elements of the matrix $\underline{C}_D$ are the expected product of the posterior means. Since the prior expectation of the posterior mean is zero, the elements are the covariances of the posterior means.

When the data is clairvoyance on the state variables $\underline{s}$, (2.4.1)

reduces to

$$\langle v_C|\mathcal{E}\rangle = -\frac{1}{2}\ \text{trace}\ \underline{E}_{co}\underline{C} \tag{2.4.7}$$

If we consider the posterior means $\langle \underline{s}|D,\mathcal{E}\rangle$ as random variables, comparison of (2.4.7) and (2.4.1) implies that the value of data is the value of clairvoyance on the posterior means. In most practical problems the value of clairvoyance on the posterior mean is much easier to compute than the value of clairvoyance on the data itself.

<u>An Interesting Special Case</u>

The most interesting special case occurs when either $\underline{E}_{co}$ or $\underline{C}_D$ is a diagonal matrix. Then the value of data becomes

$$\langle v_D|\mathcal{E}\rangle = \sum_i \underline{g}_i' \underline{H}^{-1}\underline{g}_i \overset{v}{\langle} \langle s_i|D,\mathcal{E}\rangle|\mathcal{E}\rangle \tag{2.4.8}$$

where the vector $\underline{g}_i'$ is the i^{th} row of $\underline{G}$:

$$\underline{G} = \left[\underline{g}_i'\right] \tag{2.4.9}$$

If the state variables are independent (2.4.8) is exactly equal to (2.4.1). Sufficient conditions for (2.4.8) to be a good approximation to (2.4.1) are that the diagonal elements dominate the off-diagonal elements of $\underline{E}_{co}$; that is for each i and j :

$$\rho_{ij}^2 \ll \frac{(\underline{g}_i' \underline{H}^{-1}\underline{g}_i)(\underline{g}_j' \underline{H}^{-1}\underline{g}_j)}{(\underline{g}_i' \underline{H}^{-1}\underline{g}_j)^2} \tag{2.4.10}$$

where ρ_{ij} is the correlation coefficient

$$\rho_{ij} = \langle \langle s_i|D,\mathcal{E}\rangle \langle s_j|D,\mathcal{E}\rangle|\mathcal{E}\rangle \Big/ \sqrt{\overset{v}{\langle} \langle s_i|D,\mathcal{E}\rangle|\mathcal{E}\rangle\ \overset{v}{\langle} \langle s_j|D,\mathcal{E}\rangle|\mathcal{E}\rangle}\ . \tag{2.4.11}$$

Given $\underline{G}$, $\underline{H}$, and $\underline{C}_D$, these expressions tell us when the diagonal assumption holds. A more interesting question is whether we can avoid generating the entire matrices $\underline{G}$, $\underline{H}$ and $\underline{C}_D$. The answer is yes, as shown below.

Description of the Primary Problem Using Closed Loop Sensitivities

We now show that the term $\underline{g}_i' \, \underline{H}^{-1} \underline{g}_i$ is the curvature of compensation of v with respect to the i^{th} state variable :

$$\underline{g}_i' \, \underline{H}^{-1} \underline{g}_i = \frac{\partial^2 v_{co}(s_i)}{\partial s_i^{\,2}} \tag{2.4.12}$$

where

$$v_{co}(s_i) = v_c(s_i) - v_o(s_i) \tag{2.4.13}$$

As we saw in Fig. 2.4, the open loop sensitivity is evaluated by varying s_i while the other state variables and the decision variables remain constant. We denote the open loop sensitivity as

$$v_o(s_i) = v(0,0,\ldots,s_i,\ldots,0,\underline{\hat{d}}_o) \; . \tag{2.4.14}$$

In closed loop sensitivity the state variables other than s_i remain fixed, but the decision is reoptimized for each s_i :

$$v_c(s_i) = v(0,0,\ldots,s_i,\ldots,0,\hat{d}(0,0,\ldots,s_i,\ldots,0)) \tag{2.4.15}$$

To show that expression (2.4.12) is valid we evaluate $v_o(s_i)$ and $v_c(s_i)$ for the quadratic value function (2.3.8) :

$$v_o(s_i) = a + b_i s_i + \frac{1}{2} e_{ii} s_i^{\,2} \tag{2.4.16}$$

$$v_c(s_i) = \max_{\underline{d}}(a + b_i s_i + \frac{1}{2} e_{ii} s_i^{\,2} + s_i \underline{g}_i' \underline{d} + \frac{1}{2} \underline{d}\, \underline{H}\, \underline{d})$$

$$= a + b_i s_i + \frac{1}{2} e_{ii} s_i^{\,2} - \frac{1}{2} \underline{g}_i' \underline{H}^{-1} \underline{g}_i s_i^{\,2} \qquad (2.4.17)$$

Subtracting (2.4.16) from (2.4.17) the compensation is

$$v_{co}(s_i) = -\frac{1}{2} \underline{g}_i' \underline{H}^{-1} \underline{g}_i\, s_i^{\,2} \ . \qquad (2.4.18)$$

Therefore, by evaluating the curvatures of the compensation curves for the state variables, the need to find the matrices of partial derivatives $\underline{G}$ and $\underline{H}$ is eliminated.

The Description of the Data Generating Process Through Preposterior Moments

The second component of (2.4.8) is ${}^{v}\langle\langle s_i|D,\mathcal{E}\rangle|\mathcal{E}\rangle$ the prior variance of the posterior mean. To evaluate this term we use the theorem :

$$^{v}\langle s|\mathcal{E}\rangle = {}^{v}\langle\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle + \langle{}^{v}\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle \qquad (2.4.19)$$

A proof of this theorem is given in Raiffa and Schlaifer [7, p. 106]. The theorem states that the prior variance ${}^{v}\langle s|\mathcal{E}\rangle$ has two sources. The expected posterior variance $\langle{}^{v}\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle$ is a residual variance which will not be resolved by the experiment that generates the data D . The prior variance of the posterior mean ${}^{v}\langle\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle$ is the portion of the prior variance that will be resolved by the experiment.

Sample Data

Expression (2.4.19) is best known for the case where data are N random samples from $\{s|\mathcal{E}\}$. First we consider the limiting cases of no samples and of infinite samples. Then we consider a finite number of samples.

When the data is the null experiment, $N = 0$, the prior and posterior states of information coincide. Therefore, we have

$$\langle {}^{v}\!\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle = \langle {}^{v}\!\langle s|\mathcal{E}\rangle|\mathcal{E}\rangle = {}^{v}\!\langle s|\mathcal{E}\rangle \tag{2.4.20}$$

$${}^{v}\!\langle\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle = {}^{v}\!\langle\langle s|\mathcal{E}\rangle|\mathcal{E}\rangle = 0 \; . \tag{2.4.21}$$

When the number of samples approaches infinity, the data is clairvoyance about s. The posterior probability density function will have all of its mass at a single point. Consequently, the preposterior moments are

$$\langle {}^{v}\!\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle = \langle 0|\mathcal{E}\rangle = 0 \tag{2.4.22}$$

$${}^{v}\!\langle\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle = {}^{v}\!\langle s|\mathcal{E}\rangle = {}^{v}\!\langle s|\mathcal{E}\rangle \; . \tag{2.4.23}$$

To discuss (2.4.19) for finite N it is convenient to define the ratio r

$$r = {}^{v}\!\langle\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle \big/ {}^{v}\!\langle s|\mathcal{E}\rangle \; . \tag{2.4.24}$$

The limiting cases are $r = 0$ for the null experiment and $r = 1$ for clairvoyance.

A Bayesian must assign both r and $\{s|\mathcal{E}\}$ before he can calculate the expected value of sample information. For example, Raiffa and Schlaifer [7, p. 110] suggest assigning an equivalent sample size N' to the term ${}^{v}\!\langle\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle$. Then for certain conditions the parameter r is

$$r = \frac{N}{N' + N} \; . \tag{2.4.25}$$

Assigning either r or N' weights the prior information relative to the sample information.

Experiments That Do Not Involve Sampling

Encoding and modeling are analogous to sampling because they partially resolve uncertainty about the state variable s . Encoding the parameter r or equivalently $\overset{\vee}{<}\langle s|D,\mathcal{E}\rangle|\mathcal{E}>$ should be no more difficult for these cases than for sampling.

Encoding for the Entrepreneur's Problem

Consider the demand in the Entrepreneur's Problem. One possible experiment to reduce uncertainty is to improve the deterministic model. A second possibility is to encode $\{q|\mathcal{E}\}$ more completely.

In the first case suppose the entrepreneur wants to know whether it is worthwhile to divide the market into sectors and to study historical data about consumer response to price changes. To evaluate the model improvement we encode what the new prediction might be at the price $\hat{p}_o = 24.1$. We ask questions like would you rather bet that a fair coin comes up heads on the next toss or bet that the new prediction will be within 10 percent of the original one. A series of such questions reveals that the entrepreneur's probability density function on the mean shift is normal with the mean equal to the previous estimate of 58.5 and the variance equal to 10. Since Δq was previously defined as the difference between the predicted and actual demand we assign

$$\left\langle \overset{\vee}{<} q|D,\mathcal{E}>|\mathcal{E} \right\rangle = \overset{\vee}{<} q|\mathcal{E}> = 100 \tag{2.4.26}$$

$$\overset{\vee}{<} \langle q|D,\mathcal{E}\rangle|\mathcal{E}> = 10 \ . \tag{2.4.27}$$

Consequently, using (2.4.18), the variance of demand is

$$ {}^{v}\langle q|\mathcal{E}\rangle = 110 \; . \tag{2.4.28} $$

From Section 2.2 we recall that the curvature of the compensation is

$$ \underline{E}_{co} = \frac{\partial^2 \pi_{co}(q)}{\partial q^2} = 0.095 \; . \tag{2.4.29} $$

The expected value of the modeling data D is found by substituting (2.3.27) and (2.4.29) into (2.4.1) :

$$ \langle v_D|\mathcal{E}\rangle = (0.095)(10) = 0.95 \; . \tag{2.4.30} $$

Now suppose that the entrepreneur has upgraded his model, but he has left one free parameter, q the demand at $\hat{p}_o = 24.1$. He feels that if he knew q he would have complete confidence in his model. His prior on q has moments

$$ \langle q|\mathcal{E}\rangle = 58.5 \tag{2.4.31} $$

$$ {}^{v}\langle q|\mathcal{E}\rangle = 100 \; . \tag{2.4.32} $$

Upon questioning, the entrepreneur reveals that a contributing factor to his uncertainty is personal ignorance. If he had the opportunity to incorporate his staff's expertise he is confident that $\{q|\mathcal{E}\}$ would change. After further questioning he decides that there is a 50 percent chance that he could change q by more than 5 units after learning his staff's opinion. When we point out that this implies an expected posterior variance of 50, the entrepreneur says, "That sounds reasonable." Therefore, to compute the value of encoding q we assign

$$ {}^{v}\langle q|\mathcal{E}\rangle = 100 \tag{2.4.33} $$

$$ \langle {}^{v}\langle q|D,\mathcal{E}\rangle|\mathcal{E}\rangle = 50 \tag{2.4.34} $$

$$\overset{v}{\langle} \langle q|D,\mathcal{E}\rangle|\mathcal{E}\rangle = 50 \ . \qquad (2.4.35)$$

Consequently, using (2.4.29) and (2.4.35) in (2.4.1), the value of encoding is

$$\langle v|D,\mathcal{E}\rangle = (0.095)(50) = 4.8 \ . \qquad (2.4.36)$$

2.5 The Value of Data for Non-Quadratic Decision Problems

The purpose of this section is to discuss how the theorem of Section 2.3 can be extended to non-quadratic problems. The result is a practical procedure for ranking the state variables. Given certain non-restrictive conditions the ranking scheme applies to any single-stage decision problem, regardless of whether the decision and state variables are continuous or discrete.

The Discrete Decision

Before we turn to the general evaluation scheme, we consider the discrete problem. With presubscripts denoting the decision alternative the value function is

$$v(\underline{s},d) = \begin{cases} {}_1a + {}_1\underline{b}'\underline{s} + \frac{1}{2}\underline{s}'\ {}_1\underline{E}\ \underline{s} + \ldots & d = d_1 \\ {}_2a + {}_2\underline{b}'\frac{1}{2}\underline{s}'\ {}_2\underline{E}\ \underline{s} + \ldots & d = d_2 \end{cases} \ . \qquad (2.5.1)$$

There are only two possible decisions, d_1 and d_2 .

We would like to find an expression for the value of data similar to the one derived for the quadratic problem in Section 2.3. The expression should depend on deterministic sensitivity data and the prior distribution of the posterior mean.

The key factor in the quadratic problem is that optimizing the value function evaluated at the mean of the state variables is equivalent to

optimizing the expected value :

$$\max_{\underline{d}}{}^{-1}\ v(\langle \underline{s} | D, \mathcal{E} \rangle,\ \underline{d}) = \max_{\underline{d}}{}^{-1} \langle v | \underline{d}, D, \mathcal{E} \rangle \tag{2.5.2}$$

Without this property we cannot have the simplification that the closed loop stochastic sensitivities can be replaced by deterministic sensitivities:

$$v_{co}(s_i) = \langle v | \hat{\underline{d}}(s_i, \mathcal{E}), s_i, \mathcal{E} \rangle - \langle v | \hat{d}(\mathcal{E}), \mathcal{E} \rangle \tag{2.5.3}$$

It is straightforward to show that (2.5.2) and consequently (2.5.3) hold for the discrete problem only if the value function is linear :

$$v(\underline{s}, d) = \begin{cases} {}_1a + {}_1\underline{b}\ \underline{s} & \text{if } d = d_1 \\ \\ {}_2a + {}_2\underline{b}\ \underline{s} & \text{if } d = d_2 \end{cases} \tag{2.5.4}$$

For this case consider data D_i that inpacts only s_i :

$$\langle s_j | D_i, \mathcal{E} \rangle = \langle s_j | \mathcal{E} \rangle \qquad j \neq i \tag{2.5.5}$$

We can show that the expected value of data D_i is the expected compensation for s_i :

$$\langle v_{D_i} | \mathcal{E} \rangle = \langle v_{co}(s_i) | \mathcal{E} \rangle \tag{2.5.6}$$

Figure 2.9 shows the open and closed loop deterministic sensitivities for the linear quadratic problem. The terms a and $\underline{b}$ which completely specify the closed loop sensitivity for the discrete case did not even appear in the continuous sensitivities discussed in Sections 2.2 and 2.4.

We notice that the difference between the value of data for the

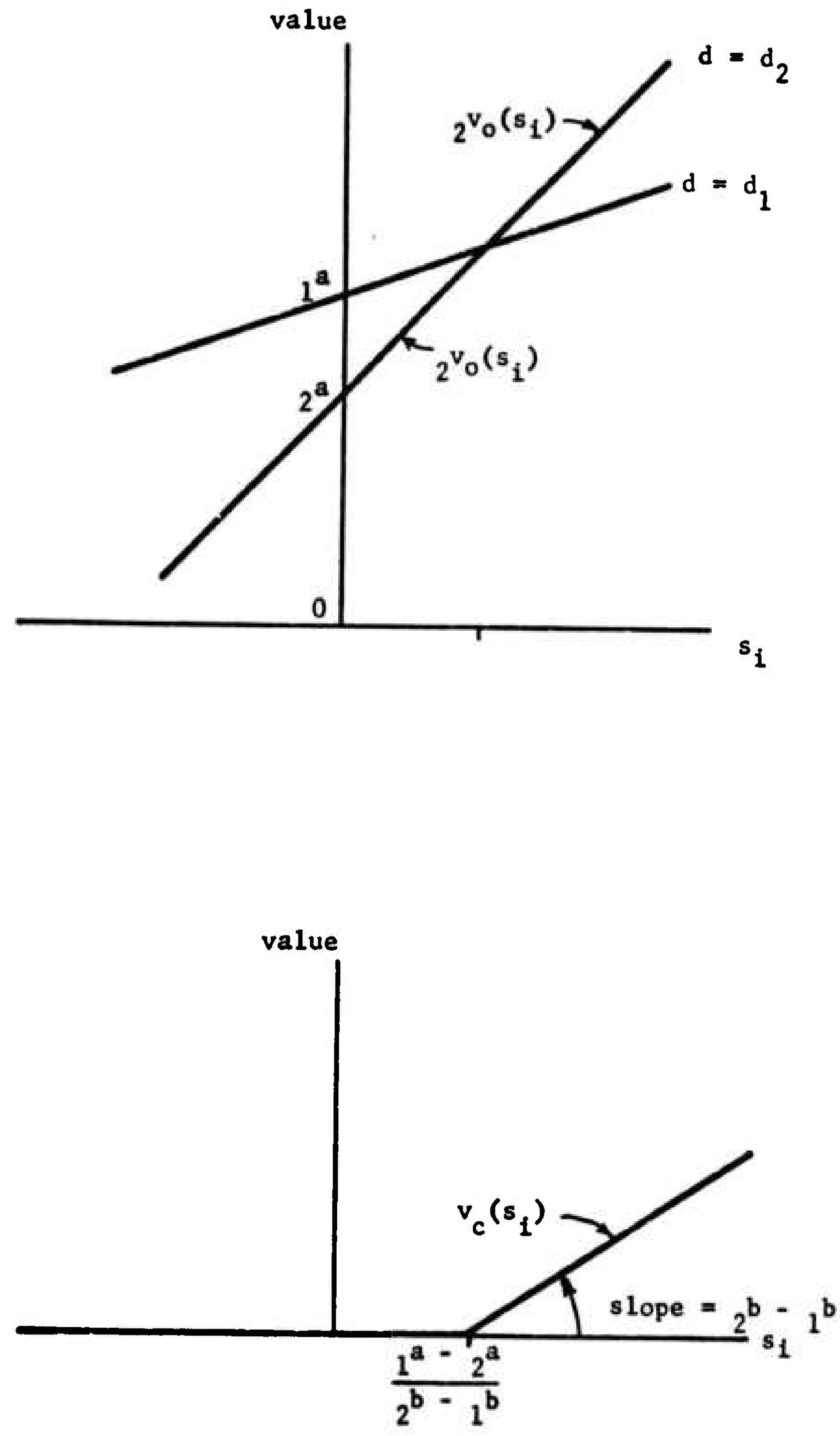


Figure 2.9 The open and closed loop sensitivities for a linear two-action problem

discrete and continuous cases is contained in the differences between the sensitivity plots. Expression (2.5.6) holds for both cases. This similarity suggests the following practical procedure for ranking state variables in a decision problem:

<u>Step 1</u>: Plot the deterministic open and closed loop sensitivities for each state variable.

<u>Step 2</u>: Calculate the expected value of compensation for each variable.

<u>Discussion of the Ranking Scheme</u>

Certain conditions must hold for the ranking to be accurate. First, the condition (2.5.3) that the stochastic compensation is approximately the deterministic compensation must hold. Second, the state variables must be effectively independent in the sense of (2.4.10) and (2.5.5). Third, if the data is not clairvoyance, the r coefficient defined in (2.4.24) must be the same for each variable, making the value of data proportional to the value of clairvoyance. For most practical problems this is not a restrictive set of assumptions.

An approximate method for performing Step 1 for the quadratic problem is suggested in the Entrepreneur's example of Section 2.2. For the discrete problem with many possible decisions the closed loop sensitivity can be found by plotting all of the open loop sensitivities and taking the maximum as a function of s_i. This technique is illustrated for the two option case in Fig. 2.9. Problems with both discrete and continuous decision variables will have more complex compensation plots than either the continuous or discrete cases.

If the compensation plot is approximately quadratic Step 2 is performed by finding the curvature of the compensation plot and multiplying

by the variance of s_i . If the compensation plot has the form of Fig. 2.9, the expected compensation is :

$$\langle v_{D_i} | \mathcal{E} \rangle = \begin{cases} |{}_2b_i - {}_1b_i| \ L^{(r)}(s_i) \\ |{}_2b_i - {}_1b_i| \ L^{(\ell)}(s_i) \end{cases} \tag{2.5.7}$$

where

$$s_b = \frac{{}_1a - {}_2a}{{}_2b_i - {}_1b_i} \tag{2.5.8}$$

and

$$L^{(r)}(s_i) = \int_{s_b}^{\infty} d\langle s_i | D_i, \mathcal{E} \rangle \ (\langle s_i | D_i, \mathcal{E} \rangle - s_b) \ \{\langle s_i | D_i, \mathcal{E} \rangle | \mathcal{E}\} \tag{2.5.9}$$

$$L^{(\ell)}(s_i) = \int_{-\infty}^{s_b} d \ \langle s_i | D_i, \mathcal{E} \rangle \ (s_b - \langle s_i | D_i, \mathcal{E} \rangle) \ \{\langle s_i | D_i, \mathcal{E} \rangle | \mathcal{E}\} \tag{2.5.10}$$

The linear loss integrals (2.5.9) and (2.5.10) are tabulated functions for the normal distribution. It is straightforward to compute the linear loss integrals for the uniform and the triangular distributions. When either the sensitivity plot or the probability density function has a complex functional form, numerical integration is required to perform Step 2.

2.6 Deliberate Introduction of Error

We deliberately introduce errors into a decision analysis if the resulting computational savings exceed the expected loss. In this section we show that the expression for loss from using the approximate probability density function $\{\underline{s} | \mathcal{E}\}^a$ instead of the accurate one $\{\underline{s} | \mathcal{E}\}$ for the quadratic problem is ℓ^a :

$$\ell^a = -\frac{1}{2} \text{ trace } \underline{E}_{co} \, \underline{S} \qquad (2.6.1)$$

where

$$\underline{S} = \langle \underline{s}' | \mathcal{E} \rangle^a \langle \underline{s} | \mathcal{E} \rangle^a \qquad (2.6.2)$$

This is the same as the value of data (2.4.1) with the covariance matrix $\underline{C}_o$ replaced by $\underline{S}$, the matrix of products of the approximate means. $\underline{S}$ is the zero matrix for the accurate probability density function $\{\underline{s} | \mathcal{E}\}$ because the state variables are normalized to have zero mean:

$$\langle \underline{s} | \mathcal{E} \rangle = \underline{0} \qquad (2.6.3)$$

To derive (2.6.1) we must distinguish between two types of error. The total change in expected value Δv^a is the difference between the maximum expected values based on the correct and approximate probability density functions :

$${}^a = \langle v | \hat{\underline{d}}(\mathcal{E}), \mathcal{E} \rangle - \langle v | \hat{\underline{d}}(\mathcal{E})^a, \mathcal{E} \rangle^a \qquad (2.6.4)$$

By adding and subtracting $\langle v | \hat{\underline{d}}(\mathcal{E})^a, \mathcal{E} \rangle$ we can divide the total change into two parts :

$$\Delta v^a = \underbrace{\langle v | \hat{\underline{d}}(\mathcal{E}), \mathcal{E} \rangle - \langle v | \hat{\underline{d}}(\mathcal{E})^a, \mathcal{E} \rangle}_{\Delta v^a_{co}} + \underbrace{\langle v | \hat{\underline{d}}(\mathcal{E})^a, \mathcal{E} \rangle - \langle v | \hat{\underline{d}}(\mathcal{E})^a, \mathcal{E} \rangle^a}_{\Delta v^a_o} \qquad (2.6.5)$$

The decomposition is illustrated in Fig. 2.10 for a single decision variable.

The loss (2.6.1) is defined as the compensation :

$$\ell^a = \Delta v^a_{co} \qquad (2.6.6)$$

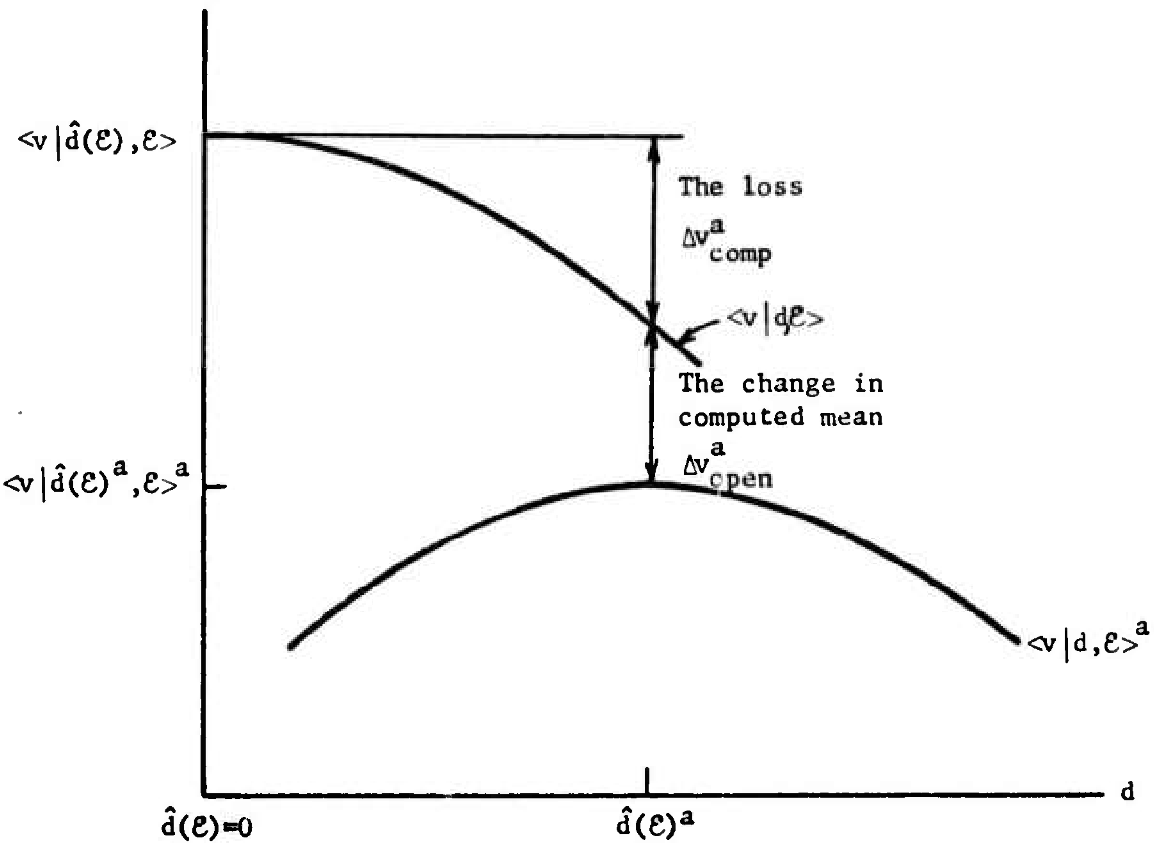


Figure 2.10 The effect of using an inaccurate probability density function on the state variables

The loss is the difference between the expected value given the best decision $\hat{\underline{d}}(\mathcal{E})$ and the expected value given the inferior decision $\hat{\underline{d}}(\mathcal{E})^a$.

To understand why Δv_o^a should not be included in the loss, suppose that the only effect of the approximation is to add the fixed amount a_o to every outcome :

$$\langle v|\underline{d},\mathcal{E}\rangle^a = \langle v|\underline{d},\mathcal{E}\rangle + a_o \tag{2.6.7}$$

Taking the gradient of both sides of (2.6.7), we find

$$\hat{\underline{d}}(\mathcal{E})^a = \hat{\underline{d}}(\mathcal{E}) \ . \tag{2.6.8}$$

Adding a_o to every outcome changes the calculated expected value without changing the decision. Using (2.6.7) and (2.6.8) in (2.6.5), the two terms are

$$\Delta v_{co}^a = 0 \tag{2.6.9}$$

$$\Delta v_o^a = - a_o \ . \tag{2.6.10}$$

Just as information only has value if it can change the decision, errors only cause losses if they affect the decision. Consequently, the expected economic loss is Δv_{co}^a , not Δv^a .

<u>Quadratic Terms</u>

We now derive expressions for Δv_{co}^a and Δv_o^a for the quadratic case. The expected value of v given $\underline{d}$ and $\{\underline{s}|\mathcal{E}\}^a$ is

$$\langle v|\underline{d},\mathcal{E}\rangle^a = a + \underline{b}'\langle\underline{s}|\mathcal{E}\rangle^a + \frac{1}{2}\langle\underline{s}'\underline{E}\ \underline{s}|\mathcal{E}\rangle^a + \langle\underline{s}|\mathcal{E}\rangle^a\ \underline{G}\ \underline{d} + \frac{1}{2}\ \underline{d}'\underline{H}\ \underline{d} \ . \tag{2.6.11}$$

Maximizing (2.6.11), we have

$$\underline{\hat{d}}(\mathcal{E})^{a} = - \underline{H}^{-1}\underline{G}' \langle\underline{s}|\mathcal{E}\rangle^{a} . \tag{2.6.12}$$

Substituting (2.6.12) into (2.6.11) yields

$$\begin{aligned} \langle v|\underline{\hat{d}}(\mathcal{E})^{a},\mathcal{E}\rangle^{a} &= a + \underline{b}'\langle\underline{s}|\mathcal{E}\rangle^{a} + \frac{1}{2} \langle\underline{s}'\underline{E}\ \underline{s}|\mathcal{E}\rangle^{a} \\ &\quad - \frac{1}{2} \langle\underline{s}'|\mathcal{E}\rangle^{a}\ \underline{G}\ \underline{H}^{-1}\underline{G}' \langle\underline{s}|\mathcal{E}\rangle^{a} . \end{aligned} \tag{2.6.13}$$

Substituting (2.6.12) into $\langle v|\underline{d},\mathcal{E}\rangle$ from (2.3.11), we have

$$\langle v|\underline{\hat{d}}(\mathcal{E})^{a},\mathcal{E}\rangle = a + \frac{1}{2} \langle\underline{s}'\underline{F}\ \underline{s}|\mathcal{E}\rangle - \frac{1}{2} \langle\underline{s}'|\mathcal{E}\rangle^{a}\ \underline{G}\ \underline{H}^{-1}\underline{G}' \langle\underline{s}|\mathcal{E}\rangle^{a} . \tag{2.6.14}$$

From (2.6.14) the prior solution is

$$\langle v|\underline{\hat{d}}(\mathcal{E}),\mathcal{E}\rangle = a + \frac{1}{2} \langle\underline{s}'\underline{E}\ \underline{s}|\mathcal{E}\rangle . \tag{2.6.15}$$

Subtracting (2.6.14) from (2.6.15) and (2.6.13) from (2.6.14) yields the desired terms :

$$\Delta v_{c}^{a} = \frac{1}{2} \langle\underline{s}'|\mathcal{E}\rangle^{a}\ \underline{G}\ \underline{H}^{-1}\underline{G}' \langle\underline{s}|\mathcal{E}\rangle^{a} \tag{2.6.16}$$

$$\Delta v_{o}^{a} = - \underline{b}' \langle\underline{s}|\mathcal{E}\rangle^{a} + \frac{1}{2} \langle\underline{s}'\underline{E}\ \underline{s}|\mathcal{E}\rangle - \frac{1}{2} \langle\underline{s}'\underline{E}\ \underline{s}|\mathcal{E}\rangle^{a} \tag{2.6.17}$$

To evaluate a computational procedure we must be careful to distinguish between open loop changes and compensatory changes. Since the open loop changes do not affect our decision, eliminating them only satisfies curiosity. It is the compensatory changes that we are willing to pay to eliminate.

CHAPTER 3

VALUE OF ANALYSIS FOR A RISK SENSITIVE DECISION MAKER

3.0 Introduction

The object of this chapter is to extend the results of Chapter 2 to a risk sensitive decision maker. We calculate the value of clairvoyance for an exponential utility function and a quadratic value function. The result is not a practical one because it involves third and fourth covariances which are difficult to encode. However, the result provides a basis for determining conditions under which the expressions from Chapter 2 are valid. In the final section of the chapter we compute the loss if risk preference is omitted from a decision analysis.

3.1 Preliminaries

The phenomenon of risk preference is well known. Pratt [5] and Howard [2] give excellent treatments of the subject. Basically, the decision maker assigns a utility function which gives a number u to every possible value v. Decision alternatives are described by the probability distribution on v or lottery that they represent. The fundamental theorem of decision theory is that one lottery is preferred to another if and only if the expected utility of the first is greater than the second. Therefore, $\underline{d}_1$ is preferred to $\underline{d}_2$ if and only if

$$\langle u(v)\,|\underline{d}_1,\mathcal{E}\rangle > \langle u(v)\,|\underline{d}_2,\mathcal{E}\rangle . \qquad (3.1.1)$$

The certain equivalent of a lottery is defined as the value $^{\sim}\langle v|\mathcal{E}\rangle$ such such that the utility of $^{\sim}\langle v|\mathcal{E}\rangle$ is equal to the expected utility of

the lottery :

$$u(\sim\langle v|\mathcal{E}\rangle) = \langle u(v)|\mathcal{E}\rangle \qquad (3.1.2)$$

Risk Aversion and Exponential Utility

Pratt [5] has shown that all of the essential information about a utility function is contained in the local risk aversion coefficient $\gamma(v)$:

$$\gamma(v) = -\frac{d^2u(v)}{dv^2} \Big/ \frac{du(v)}{dv} \qquad (3.1.3)$$

A constant risk aversion coefficient γ implies the exponential utility curve

$$u(v) = \frac{1 - e^{-\gamma v}}{1 - e^{-\gamma}} \quad . \qquad (3.1.4)$$

The local risk aversion for other utility functions can be conveniently represented by a power series about the mean of the lottery $\langle v|\mathcal{E}\rangle$:

$$\gamma(v) = \gamma(\langle v|\mathcal{E}\rangle) + \frac{d\gamma(\langle v|\mathcal{E}\rangle)}{dv}(v - \langle v|\mathcal{E}\rangle) + \ldots \qquad (3.1.5)$$

From (3.1.5) we see that any utility curve can be approximated by the exponential utility curve with

$$\gamma = \gamma(\langle v|\mathcal{E}\rangle) \qquad (3.1.6)$$

as long as the variance of v is not too large.

The exponential utility function is convenient analytically because it has the delta property; if a fixed amount δ is added to each of the prizes in a lottery, then the certain equivalent is also increased by δ :

$$^{\sim}\langle v + \delta|\mathcal{E}\rangle = {}^{\sim}\langle v|\mathcal{E}\rangle + \delta \tag{3.1.7}$$

The Approximate Certain Equivalent

An approximation for the certain equivalent is

$$^{\sim}\langle v|\mathcal{E}\rangle = \langle v|\mathcal{E}\rangle - \frac{1}{2}\,\gamma(\langle v|\mathcal{E}\rangle)\,{}^{v}\langle v|\mathcal{E}\rangle\ . \tag{3.1.8}$$

Expression (3.1.8) is exact for exponential utility and a normal lottery. Therefore, for lotteries which are approximately symmetric and not too diffuse, (3.1.8) should be an excellent approximation.

The Risk-Sensitive Value of Clairvoyance

The risk-sensitive value of clairvoyance is defined as the cost k such that the expected utilities with and without clairvoyance are equal :

$$\langle u|k,C,\mathcal{E}\rangle = \langle u|\mathcal{E}\rangle \tag{3.1.9}$$

Using the delta property and the definition of clairvoyance, it is straightforward to show that for exponential utility (3.1.9) reduces to

$$k = {}^{\sim}\langle v_c|\mathcal{E}\rangle = {}^{\sim}\langle v|\underline{d}^*(C,\mathcal{E}),\mathcal{E}\rangle - {}^{\sim}\langle v|\hat{\underline{d}}(\mathcal{E}),\mathcal{E}\rangle\ . \tag{3.1.10}$$

The second term in (3.1.10) is the risk-sensitive solution to the basic problem. In this chapter $\hat{\underline{d}}(\mathcal{E})$ represents

$$\hat{\underline{d}}(\mathcal{E}) = \max_{\underline{d}}{}^{-1}\ {}^{\sim}\langle v|\underline{d},\mathcal{E}\rangle\ . \tag{3.1.11}$$

When we wish to refer to the $\hat{\underline{d}}(\mathcal{E})$ of Chapter 2 that maximizes the expected value of v we will use $\hat{\underline{d}}_o$ since we showed in Chapter 2 that $\hat{\underline{d}}_o$ maximizes $\langle v|\underline{d},\mathcal{E}\rangle$ as well as $v(\langle \underline{s}|\mathcal{E}\rangle,\underline{d})$.

The first term in (3.1.10) may be simplified. Clairvoyance allows

us to maximize $\underline{d}$ after $\underline{s}$ is revealed :

$$\hat{\underline{d}}(C,\mathcal{E}) = \hat{\underline{d}}(\underline{s}) \qquad (3.1.13)$$

Given both $\underline{s}$ and $\underline{d}$, v is known deterministically; therefore, we have

$$\hat{\underline{d}}(\underline{s}) = \max_{\underline{d}}{}^{-1} \sim\langle v|\underline{s},\underline{d},\mathcal{E}\rangle = \max_{\underline{d}}{}^{-1} v(\underline{s},\underline{d}) \ . \qquad (3.1.14)$$

Since the resulting certain equivalent is a function only of $\underline{s}$, we may write

$$\sim\langle v|\underline{d}^*(C,\mathcal{E}),\mathcal{E}\rangle = \sim\langle v(\underline{s},\hat{\underline{d}}(s))|\mathcal{E}\rangle \ . \qquad (3.1.15)$$

To find the certain equivalent given clairvoyance, we maximize the determistic value function, perform a change of variables from $\underline{s}$ to v , and find the certain equivalent of the resulting lottery.

3.2 The Value of Clairvoyance for a Risk-Sensitive Quadratic Problem

Although it is possible to definte the value of data for a risk-sensitive decision maker, the resulting expressions are so complex that we gain little insight. Instead, we examine the special case where the data is clairvoyance on the vector of state variables. By examining the conditions under which the risk-sensitive value of clairvoyance reduces to the risk-indifferent value of clairvoyance, we learn when we can apply the results from Chapter 2.

The Derivation of the Approximate Value of Clairvoyance

To find the value of clairvoyance (3.2.1), we know from (3.1.10) that we find the difference between the certain equivalent with clairvoyance and the certain equivalent without.

Starting with the primary problem, we want to find

$$\sim\langle v \,|\, \underline{\hat{d}}(\mathcal{E}), \mathcal{E}\rangle^{a} = \langle v \,|\, \underline{\hat{d}}(\mathcal{E}), \mathcal{E}\rangle - \frac{1}{2}\,\gamma \sim\langle v \,|\, \underline{\hat{d}}(\mathcal{E}), \mathcal{E}\rangle \,. \qquad (3.2.2)$$

We have used the small symbol a to denote an approximation. We define $\underline{\hat{d}}(\mathcal{E})$ as

$$\underline{\hat{d}}(\mathcal{E})^{a} = \max_{\underline{d}}{}^{-1} \sim\langle v \,|\, \underline{d}, \mathcal{E}\rangle^{a} \,. \qquad (3.2.3)$$

Then our approximation is

$$\sim\langle v \,|\, \underline{\hat{d}}(\mathcal{E}), \mathcal{E}\rangle \cong \sim\langle v \,|\, \underline{\hat{d}}(\mathcal{E})^{a}, \mathcal{E}\rangle^{a} \,. \qquad (3.2.4)$$

To find $\sim\langle v \,|\, \underline{d}, \mathcal{E}\rangle^{a}$ we need the mean and variance. From (2.3.11) the mean is

$$\langle v \,|\, \underline{d}, \mathcal{E}\rangle = a + \frac{1}{2}\langle \underline{s}'\underline{E}\,\underline{s} \,|\, \mathcal{E}\rangle + \frac{1}{2}\,\underline{d}'\underline{H}\,\underline{d} \,. \qquad (3.2.5)$$

To find the variance we square expression (2.3.8) for $v(\underline{s},\underline{d})$:

$$\begin{aligned} v^{2}(\underline{s},\underline{d}) = {} & a + \underline{b}'\underline{s}\,\underline{s}'\underline{b} + \frac{1}{4}(\underline{s}'\underline{E}\,\underline{s})^{2} + \underline{d}'\underline{G}'\underline{s}\,\underline{s}'\underline{G}\,\underline{d} + \frac{1}{4}(\underline{d}'\underline{H}\,\underline{d})^{2} \\ & + 2a\,\underline{b}'\underline{s} + a\,\underline{s}'\underline{E}\,\underline{s} + 2a\,\underline{s}'\underline{G}\,\underline{d} + a\,\underline{d}'\underline{H}\,\underline{d} + \underline{b}'\underline{s}\,\underline{s}'\underline{E}\,\underline{s} \\ & + 2\underline{b}'\underline{s}\,\underline{s}'\underline{G}\,\underline{d} + \underline{b}'\underline{s}\,\underline{d}'\underline{H}\,\underline{d} + \underline{s}'\underline{E}\,\underline{s}\,\underline{s}'\underline{G}\,\underline{d} + \frac{1}{2}\,\underline{s}'\underline{E}\,\underline{s}\,\underline{d}'\underline{H}\,\underline{d} \\ & + \underline{s}'\underline{G}\,\underline{d}\,\underline{d}'\underline{H}\,\underline{d} \end{aligned} \qquad (3.2.6)$$

Abbreviating the mean vector and the covariance matrix,

$$\underline{\bar{s}} = \langle \underline{s} \,|\, \mathcal{E}\rangle \qquad (3.2.7)$$

$$\underline{C} = \langle \underline{s}\,\underline{s}' \,|\, \mathcal{E}\rangle \,, \qquad (3.2.8)$$

the expectation of (3.2.6) is

$$\langle v^2|\underline{d},\mathcal{E}\rangle = a + \underline{b}'\underline{C}\,\underline{b} + \frac{1}{4}\langle(\underline{s}'\underline{E}\,\underline{s})^2|\mathcal{E}\rangle + \underline{d}'\underline{G}'\underline{C}\,\underline{G}\,\underline{d} + \frac{1}{4}(\underline{d}'\underline{H}\,\underline{d})^2$$

$$+\, a\,\langle\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle + a\,\underline{d}'\underline{H}\,\underline{d} + \underline{b}'\langle\underline{s}\,\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle + 2\underline{b}'\underline{C}\,\underline{G}\,\underline{d}$$

$$+\,\langle\underline{s}'\underline{E}\,\underline{s}\,\underline{s}'|\mathcal{E}\rangle\,\underline{G}\,\underline{d} + \frac{1}{2}\langle\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle\,\underline{d}'\underline{H}\,\underline{d}\;. \qquad (3.2.9)$$

Squaring $\langle v|\underline{d},\mathcal{E}\rangle$ from (3.2.5) we have

$$\langle v|\underline{d},\mathcal{E}\rangle^2 = a + \frac{1}{4}\langle\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle^2 + \frac{1}{4}(\underline{d}'\underline{H}\,\underline{d})^2 + a\,\langle\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle$$

$$+\, a\,\underline{d}'\underline{H}\,\underline{d} + \frac{1}{2}\langle\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle\,\underline{d}'\underline{H}\,\underline{d}\;. \qquad (3.2.10)$$

Subtracting (3.2.10) from (3.2.9) we calculate the variance :

$${}^{v}\langle v|\underline{d},\mathcal{E}\rangle = \underline{b}'\underline{C}\,\underline{b} + \frac{1}{4}\left(\langle(\underline{s}'\underline{E}\,\underline{s})^2|\mathcal{E}\rangle - \langle\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle^2\right) + \underline{d}'\underline{G}'\underline{C}\,\underline{G}\,\underline{d}$$

$$+\,\underline{b}'\,\langle\underline{s}\,\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle + 2\underline{b}'\underline{C}\,\underline{G}\,\underline{d} + \langle\underline{s}'\underline{E}\,\underline{s}\,\underline{s}'|\mathcal{E}\rangle\,\underline{G}\,\underline{d} \qquad (3.2.11)$$

Combining (3.2.5) and (3.2.11) using (3.2.2), we have

$${}^{\sim}\langle v|\underline{d},\mathcal{E}\rangle^{a} = a + \frac{1}{2}\langle\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle - \frac{1}{2}\gamma\left(\underline{b}'\underline{C}\,\underline{b} + \frac{1}{4}\left(\langle(\underline{s}'\underline{E}\,\underline{s})^2|\mathcal{E}\rangle\right.\right.$$

$$\left.\left. -\,\langle\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle^2\right) + \underline{b}'\langle\underline{s}\,\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle\right)$$

$$-\,\frac{1}{2}\gamma\,(2\underline{b}'\underline{C} + \langle\underline{s}'\underline{E}\,\underline{s}\,\underline{s}'|\mathcal{E}\rangle)\,\underline{G}\,\underline{d} + \frac{1}{2}\underline{d}'(\underline{H} - \gamma\,\underline{G}'\underline{C}\,\underline{G})\,\underline{d}\;.$$

$$(3.2.12)$$

The gradient is

$$\nabla\,{}^{\sim}\langle v|\underline{d},\mathcal{E}\rangle^{a} = -\gamma\,(\underline{b}'\underline{C} + \frac{1}{2}\langle\underline{s}'\underline{E}\,\underline{s}\,\underline{s}'|\mathcal{E}\rangle)\,\underline{G} + \underline{d}'(\underline{H} - \gamma\,\underline{G}'\underline{C}\,\underline{G})\;.$$

$$(3.2.13)$$

Setting the gradient to zero, we find

$$\hat{\underline{d}}(\mathcal{E})^{a} = \gamma(\underline{H} - \gamma\,\underline{G}'\underline{C}\,\underline{G})^{-1}\underline{G}'(\underline{C}\,\underline{b} + \frac{1}{2}\langle\underline{s}\,\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle)\;. \qquad (3.2.14)$$

Finally, substituting (3.2.14) into (3.2. 12), the solution is

$$\begin{aligned}\sim\langle v|\hat{\underline{d}}(\mathcal{E})^{a},\mathcal{E}\rangle^{a} = a &+ \frac{1}{2}\langle\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle - \frac{1}{2}\gamma\Big(\underline{b}'\underline{C}\,\underline{b} + \frac{1}{4}\Big(\langle(\underline{s}'\underline{E}\,\underline{s})^{2}|\mathcal{E}\rangle \\ &- \langle\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle^{2}\Big) - \underline{b}'\langle\underline{s}\,\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle\Big) \\ &- \frac{1}{2}\gamma^{2}(\underline{b}'\underline{C} + \frac{1}{2}\langle\underline{s}'\underline{E}\,\underline{s}\,\underline{s}'|\mathcal{E}\rangle)\,\underline{G}(\underline{H} \\ &- \gamma\,\underline{G}'\underline{C}\,\underline{G})^{-1}\,\underline{G}'(\underline{C}\,\underline{b} + \frac{1}{2}\langle\underline{s}\,\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle)\ . \end{aligned} \qquad (3.2.15)$$

We now consider the other term in (3.1.10), the certain equivalent with clairvoyance. Expression (3.1.14) indicates that we should maximize the value function given $\underline{s}$. The gradient of $v(\underline{s},\underline{d})$ is

$$\nabla v(\underline{s},\underline{d}) = \underline{s}'\underline{G} + \underline{d}'\underline{H}\ . \qquad (3.2.16)$$

Setting the gradient to zero yields

$$\hat{\underline{d}}(\underline{s}) = -\,\underline{H}^{-1}\underline{G}'\underline{s}\ . \qquad (3.2.17)$$

Notice that this is the exact maximum since we have not yet introduced the certain equivalent approximation for this case. Substituting into (2.3.8), we have

$$v(\underline{s},\hat{\underline{d}}(\underline{s})) = a + \underline{b}'\underline{s} + \frac{1}{2}\,\underline{s}'(\underline{E} - \underline{G}\,\underline{H}^{-1}\underline{G}')\,\underline{s}\ . \qquad (3.2.18)$$

Proceeding in analogy to (3.2.5) through (3.2.11), the mean and variance are

$$\langle v(\underline{s},\hat{\underline{d}}(\underline{s}))|\mathcal{E}\rangle = a + \frac{1}{2}\langle\underline{s}'(\underline{E} - \underline{G}'\underline{H}^{-1}\underline{G}')\,\underline{s}|\mathcal{E}\rangle\ , \qquad (3.2.19)$$

$$\begin{aligned}{}^{v}\langle v(\underline{s},\hat{\underline{d}}(\underline{s}))|\mathcal{E}\rangle = \underline{b}'\underline{C}\,\underline{b} &+ \frac{1}{4}\Big(\langle(\underline{s}'(\underline{E} - \underline{G}\,\underline{H}^{-1}\underline{G}')\,\underline{s})^{2}|\mathcal{E}\rangle - \langle\underline{s}'(\underline{E} \\ &- \underline{G}\,\underline{H}^{-1}\underline{G}')\underline{s}|\mathcal{E}\rangle^{2}\Big) + \underline{b}'\langle\underline{s}\,\underline{s}'(\underline{E} - \underline{G}\,\underline{H}^{-1}\underline{G}')\,\underline{s}|\mathcal{E}\rangle\ . \end{aligned} \qquad (3.2.20)$$

Combining (3.2.19) and (3.2.20), using (3.2.2), we have

$$\tilde{\langle} v(\underline{s},\underline{\hat{d}}(\underline{s}))\,|\mathcal{E}\rangle^{a} = a + \frac{1}{2}\,\langle\underline{s}'(\underline{E} - \underline{G}\;\underline{H}^{-1}\underline{G}')\underline{s}\,|\mathcal{E}\rangle - \frac{1}{2}\,\gamma\,\Big(\underline{b}'\underline{C}\;\underline{b}$$

$$+ \frac{1}{4}\,\Big(\langle(\underline{s}'(\underline{E} - \underline{G}\;\underline{H}^{-1}\underline{G}')\underline{s})^{2}\,|\mathcal{E}\rangle - \langle\underline{s}'(\underline{E} - \underline{G}\;\underline{H}^{-1}\underline{G}')\underline{s}\,|\mathcal{E}\rangle^{2}\Big)$$

$$+ \underline{b}'\langle\underline{s}\;\underline{s}'(\underline{E} - \underline{G}\;\underline{H}^{-1}\underline{G}')\underline{s}\,|\mathcal{E}\rangle\Big)\,. \qquad (3.2.21)$$

Finally, subtracting (3.2.15) from (3.2.21) we have the result :

$$\tilde{\langle} v_{c}\,|\mathcal{E}\rangle^{a} = -\,\frac{1}{2}\,\langle\underline{s}'\underline{G}\;\underline{H}^{-1}\underline{G}'\underline{s}\,|\mathcal{E}\rangle - \frac{1}{8}\,\gamma\,\Big(\langle(\underline{s}'\underline{G}\;\underline{H}^{-1}\underline{G}'\underline{s})^{2}\,|\mathcal{E}\rangle$$

$$-\ \langle\underline{s}'\underline{G}\;\underline{H}^{-1}\underline{G}'\underline{s}\,|\mathcal{E}\rangle^{2} - 2\,\langle\underline{s}'\underline{G}\;\underline{H}^{-1}\underline{G}'\underline{s}\;\underline{s}'\underline{E}\;\underline{s}\,|\mathcal{E}\rangle$$

$$+ 2\,\langle\underline{s}'\underline{G}\;\underline{H}^{-1}\underline{G}'\underline{s}\;\underline{s}'\underline{E}\;\underline{s}\,|\mathcal{E}\rangle^{2} + \underline{b}'\,\langle\underline{s}\;\underline{s}'\underline{G}\;\underline{H}^{-1}\underline{G}'\underline{s}\,|\mathcal{E}\rangle\Big)$$

$$+ \frac{1}{2}\,\gamma\,\Big(\underline{b}'\underline{C} + \frac{1}{2}\,\langle\underline{s}'\underline{E}\;\underline{s}\;\underline{s}'\,|\mathcal{E}\rangle\Big)\,\underline{G}(\underline{H} - \gamma\,\underline{G}'\underline{C}\;\underline{G})^{-1}\underline{G}'(\underline{C}\;\underline{b}$$

$$+ \frac{1}{2}\,\langle\underline{s}\;\underline{s}'\underline{E}\;\underline{s}\,|\mathcal{E}\rangle\Big) \qquad (3.2.22)$$

Discussion of the Risk-Sensitive Value of Clairvoyance

We now examine conditions under which the risk indifferent value of clairvoyance (2.4.7) is a good approximation to the risk-sensitive value of clairvoyance (3.2.22). Since for γ equal to zero the two expressions must be equal, the first term of (3.2.22) is the risk-indifferent value of clairvoyance :

$$\langle v_{c}\,|\mathcal{E}\rangle = -\,\frac{1}{2}\,\langle\underline{s}'\underline{G}\;\underline{H}^{-1}\underline{G}'\underline{s}\,|\mathcal{E}\rangle \qquad (3.2.23)$$

Therefore, we want conditions such that the first term of (3.2.22) dominates the others.

To eliminate third and fourth covariances, we assume that the state variables are normal and independent. Under these assumptions

the following properties are true :

$$\mathrm{cov}\langle s_i, s_j | \mathcal{E}\rangle = \begin{cases} \overset{\vee}{s}_i & i=j \\ 0 & i\neq j \end{cases} \qquad (3.2.24)$$

$$\mathrm{cov}\langle s_i, s_j, s_k | \mathcal{E}\rangle = 0 \qquad \text{all } i,j,k \qquad (3.2.25)$$

$$\mathrm{cov}\langle s_i, s_j, s_k, s_\ell | \mathcal{E}\rangle = \begin{cases} 3\,\overset{\vee}{s}{}_i^2 & i=j=k=\ell \\ \overset{\vee}{s}_i\overset{\vee}{s}_k & i=j,\ k=\ell \\ \overset{\vee}{s}_i\overset{\vee}{s}_j & i=k,\ j=\ell \text{ or } i=\ell,\ j=k \\ 0 & \text{otherwise} \end{cases} \qquad (3.2.26)$$

We use the abbreviated symbol for variance $\overset{\vee}{s}$, when the state of information is understood to be $\mathcal{E}$. Using (3.2.25) to eliminate the third covariances, sufficient conditions for the first term of (3.2.22) to dominate the others are :

$$\langle v_c | \mathcal{E}\rangle \gg \frac{1}{8}\,\gamma\left(\langle(\underline{s}'\underline{E}_{co}\underline{s})^2|\mathcal{E}\rangle - \langle\underline{s}'\underline{E}_{co}\underline{s}|\mathcal{E}\rangle^2\right) \qquad (3.2.27)$$

$$\langle v_c | \mathcal{E}\rangle \gg \frac{1}{4}\,\gamma\left|\langle\underline{s}'\underline{E}_{co}\underline{s}\ \underline{s}'\underline{E}\ \underline{s}|\mathcal{E}\rangle - \langle\underline{s}'\underline{E}_{co}\underline{s}|\mathcal{E}\rangle\langle\underline{s}'\underline{E}\ \underline{s}|\mathcal{E}\rangle\right| \qquad (3.2.28)$$

$$\langle v_c | \mathcal{E}\rangle \gg \frac{1}{2}\,\gamma^2\ \underline{b}'\underline{C}\ \underline{G}(\underline{H} - \gamma\,\underline{G}'\underline{C}\ \underline{G})^{-1}\underline{G}'\underline{C}\ \underline{b} \qquad (3.2.29)$$

where we have used the definition of $\underline{E}_{co}$ from Chapter 2 .

$$\underline{E}_{co} = \underline{G}\ \underline{H}^{-1}\underline{G}' \qquad (3.2.30)$$

Applying (3.2.24) and (3.2.26), the conditions are simplified :

$$\frac{\langle v_c | \mathcal{E}\rangle}{\gamma} \gg \frac{1}{4}\sum_i\sum_j e^2_{co_{ij}}\,\overset{\vee}{s}_i\overset{\vee}{s}_j \qquad (3.2.30)$$

$$\frac{\langle v_c|\mathcal{E}\rangle}{\gamma} \gg \frac{1}{2}\sum_i\sum_j e_{co_{ij}}\; e_{ij}\; \overset{v}{s}_i \overset{v}{s}_j \tag{3.2.31}$$

$$\frac{\langle v_c|\mathcal{E}\rangle}{\gamma^2} \gg \underline{b}'\underline{C}\;\underline{G}(\underline{H} - \gamma\,\underline{G}'\underline{C}\;\underline{G})^{-1}\underline{G}'\underline{C}\;\underline{b} \tag{3.2.32}$$

Expressions (3.2.30) through (3.2.32) depend only on the first and second order partial derivatives of v and on the covariances of the state variables.

<u>One State Variable and One Decision Variable</u>

To better understand the three conditions (3.2.30), (3.2.31), and (3.2.32) we specialize them to the case of one state variable and one decision variable. Recognizing that the value of clairvoyance may be written,

$$\langle v_c|\mathcal{E}\rangle = -\frac{1}{2}\, e_{co}\, \overset{v}{s}\ , \tag{3.2.33}$$

the conditions reduce to

$$\frac{1}{\gamma} \gg \langle v_c|\mathcal{E}\rangle \tag{3.2.34}$$

$$\frac{1}{\gamma} \gg |e\ \overset{v}{s}| \tag{3.2.35}$$

$$\frac{1}{\gamma^2} \gg \frac{b^2 \overset{v}{s}}{1 + 2\gamma\langle v_c|\mathcal{E}\rangle} \tag{3.2.36}$$

The second two conditions, (3.2.35) and (3.2.36), are no surprise. Assuming (3.2.34) holds, (3.2.35) and (3.2.36) together imply

$$\frac{1}{\gamma^2} \gg b^2\, \overset{v}{s} + \frac{1}{2}\, e^2\, \overset{v}{s}{}^2 \approx {}^v\langle v|d=0,\mathcal{E}\rangle\ . \tag{3.2.37}$$

The criterion that the variance be much less than the reciprocal of the risk aversion coefficient is the one usually invoked to justify the certain equivalent approximation (3.1.8) (see Howard [2, p. 513]).

The new insight is that the value of clairvoyance on the state variable must be small relative to the reciprocal of the risk aversion coefficient. Even if the variance criterion (3.2.37) is satisfied, a problem is unsuitable for approximate analysis if strong coupling between the state and decision variables causes a violation of (3.2.34).

3.3 The Approximate Value of Risk Preference

In this section we consider both the loss from deliberate suppression of risk preference and the gain from additional assessment. These are important quantities for the applications of Chapter 5.

The most interesting result of this section is that the risk aversion coefficient γ can be treated as a random variable. The approximate value of risk preference encoding is the approximate value of information on γ, treating γ as if it were a state variable.

Deliberate Suppression of Risk Preference

Deliberate suppression of risk preference results in an inaccurate decision. The loss from using the risk-indifferent optimum $\hat{\underline{d}}_o$ instead of the risk-sensitive optimum $\hat{\underline{d}}(\mathcal{E})$ is

$$\ell_\gamma = \sim\langle v | \hat{\underline{d}}(\mathcal{E}), \mathcal{E}\rangle - \sim\langle v | \hat{\underline{d}}_o, \mathcal{E}\rangle . \qquad (3.3.1)$$

For a single decision variable, (3.3.1) is illustrated by the lower curve in Fig. 3.1.

As discussed in the previous section the exact certain equivalent

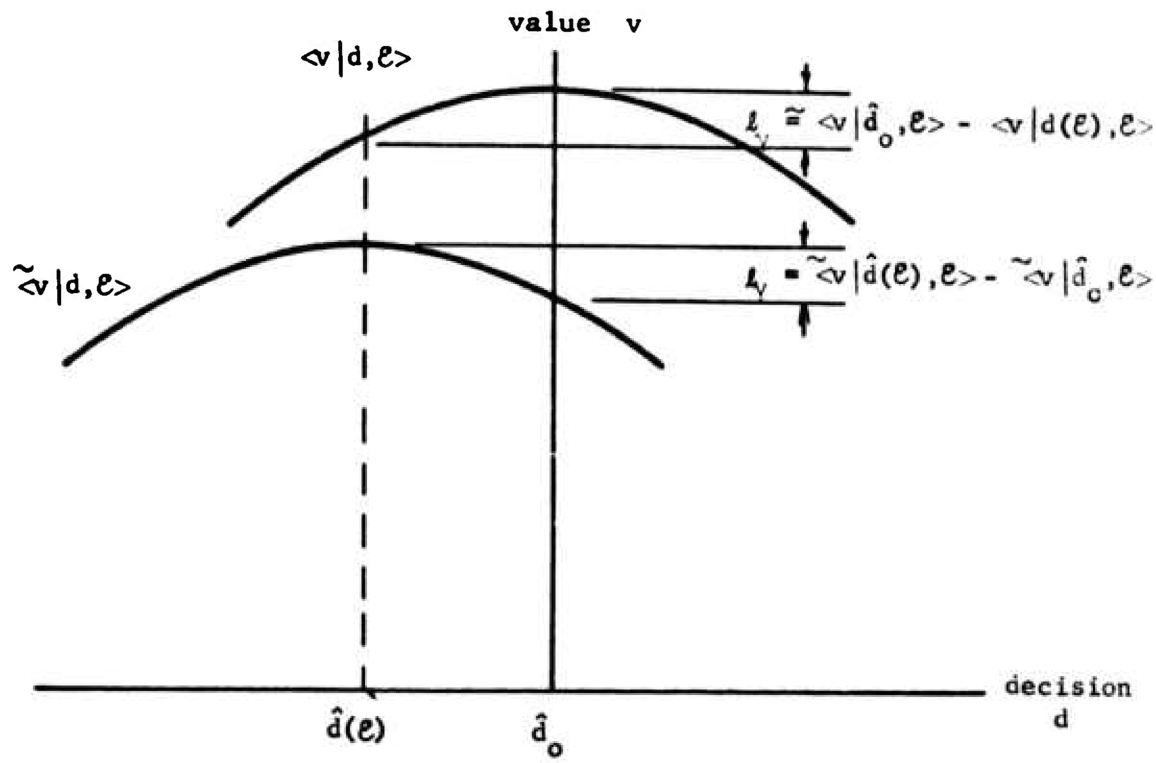


Figure 3.1 The exact and approximate losses from suppression of risk preference

is difficult to compute. Therefore, we approximate (3.3.1) by

$$\ell_\gamma^a = {}^\sim v\langle\hat{\underline{d}}(\mathcal{E})^a,\mathcal{E}\rangle^a - {}^\sim v\langle\hat{\underline{d}}_o,\mathcal{E}\rangle^a , \tag{3.3.2}$$

where the superscript a denotes the certain equivalent approximation (3.1.8). The first term in (3.3.2) is (3.2.15) and the second term is (3.2.12) evaluated at $\hat{\underline{d}}_o = \underline{0}$. Consequently, we have

$$\ell_\gamma^a = -\frac{1}{2}\gamma^2(\underline{b}'\underline{C} + \frac{1}{2}\langle\underline{s}'\underline{E}\,\underline{s}\,\underline{s}'|\mathcal{E}\rangle)\,\underline{G}(\underline{H} - \gamma\,\underline{G}'\underline{C}\,\underline{G})^{-1}\underline{G}'$$

$$(\underline{C}\,\underline{b} + \frac{1}{2}\langle\underline{s}\,\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle) . \tag{3.3.3}$$

If we were only interested in the suppression of risk aversion (3.3.3) would suffice. However, for additional risk preference assessment, the expression analogous to (3.3.3) is difficult to derive. Therefore, we now show that a second approximation of ℓ_γ is

$$\ell_\gamma^{am} = v\langle\hat{\underline{d}}_o,\mathcal{E}\rangle - v\langle\hat{\underline{d}}(\mathcal{E})^a,\mathcal{E}\rangle \tag{3.3.4}$$

where the superscript am denotes the approximation based on the mean. The approximation is illustrated in the top curve of Fig. 3.1.

To motivate the approximation we return to the case of one state variable and one decision variable. We assume that condition (3.2.34) holds; the reciprocal of the risk aversion coefficient is large compared to the value of clairvoyance on the state variable. Using the definition of clairvoyance, this condition implies that

$$-\frac{\partial^2 v}{\partial d^2}\bigg|_o \gg \frac{1}{2}\gamma\left(\frac{\partial^2 v}{\partial s\,\partial d}\bigg|_o\right)^2 {}^{v}\langle s|\mathcal{E}\rangle . \tag{3.3.5}$$

Differentiating the expressions for the mean and variance, (3.2.5) and

(3.2.11), we see that (3.3.5) implies

$$- \left. \frac{\partial^2 \langle v | d, \mathcal{E} \rangle}{\partial d^2} \right|_o \gg \frac{1}{2} \gamma \left. \frac{\partial^2 {}^v\langle v | d, \mathcal{E} \rangle}{\partial d^2} \right|_o . \tag{3.3.6}$$

Therefore, when condition (3.2.34) holds the variance is approximately linear in d .

In Fig. 3.2 we show the relationship between the mean and certain equivalent when the variance is linear in d . Using (3.1.8), the certain equivalent is the mean minus the risk premium, one half γ times the variance. At $\hat{d}_o$ the slope of the expected value is zero; therefore, the slope of the certain equivalent at $\hat{d}_o$ is minus the slope of the risk premium. Likewise, at $\hat{d}(\mathcal{E})^a$ the slope of the certain equivalent is zero, making the slope of the mean equal to the slope of the risk premium. By (3.2.5) and (3.2.11) both the mean and variance are quadratic in d . Since the risk premium is linear in d , both the mean and certain equivalent must have the same constant second partial derivative with respect to d .

The conclusion is that the curves for the mean and certain equivalent in Fig. 3.2 are identical parabolas; if we translated the curve for the certain equivalent to the right by

$$\Delta d = \hat{d}_o - \hat{d}(\mathcal{E})^a \tag{3.3.7}$$

and upward by

$$\Delta v = \langle v | \hat{d}_o, \mathcal{E} \rangle - \langle\!\sim v | \hat{d}(\mathcal{E})^a, \mathcal{E} \rangle^a , \tag{3.3.8}$$

the two curves would coincide. It follows that the two approximations (3.3.2) and (3.3.4) are the same for this case :

$$\ell_\gamma^{am} = \ell_\gamma^a \tag{3.3.9}$$

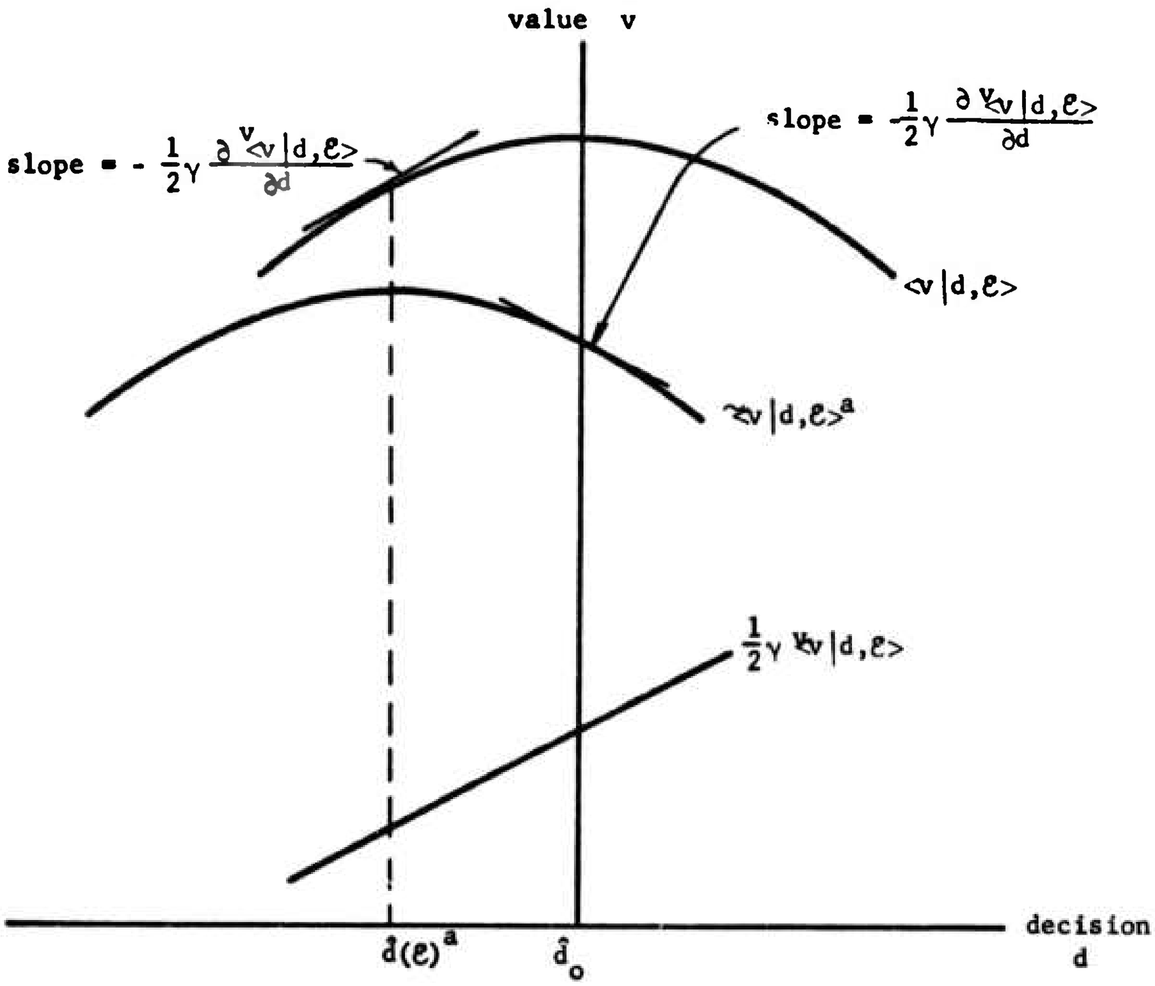


Figure 3.2 Relationship between the expected value and certain equivalent when the variance is a linear function of the decision

The Value of Additional Assessment

To evaluate the value of additional risk assessment we consider γ as a random variable. Assuming that γ is the only useful data from risk preference encoding, we would like to compute

$$\langle v_{\gamma}|\mathcal{E}\rangle = \sim\!v\lfloor \underline{d}^{*}(\gamma,\mathcal{E}),\mathcal{E}\rangle - \sim\!v\lfloor \hat{\underline{d}}(\mathcal{E}),\mathcal{E}\rangle \,. \qquad (3.3.10)$$

Unfortunately, the first term in (3.3.10) is difficult to calculate since the expansion rule does not hold for certain equivalents:

$$\sim\!v\lfloor \underline{d}^{*}(\gamma,\mathcal{E}),\mathcal{E}\rangle \neq \widetilde{\langle}\langle v\lfloor \hat{\underline{d}}(\gamma,\mathcal{E}),\gamma,\mathcal{E}\rangle|\mathcal{E}\rangle \qquad (3.3.11)$$

Consequently, we define the approximate expected gain :

$$\begin{aligned}\langle v_{\gamma}|\mathcal{E}\rangle^{am} &= \langle v\lfloor \hat{\underline{d}}(\mathcal{E})^{a},\mathcal{E}\rangle - \langle v\lfloor \underline{d}^{*}(\gamma,\mathcal{E})^{a},\mathcal{E}\rangle \\ &= \langle v\lfloor \hat{\underline{d}}(\mathcal{E})^{a},\mathcal{E}\rangle - \langle\langle v\lfloor \hat{\underline{d}}(\gamma,\mathcal{E})^{a},\gamma,\mathcal{E}\rangle|\mathcal{E}\rangle \qquad (3.3.12)\end{aligned}$$

The approximation is good when the analog to (3.3.5) holds; the Hessian of the mean $\underline{H}$ should dominate the Hessian of the risk premium $-\gamma\underline{G}'\underline{C}\,\underline{G}$. The optimal decisions in (3.3.12) are approximate because they are based on the certain equivalent approximation (3.1.8) and because we assume that the solution prior to encoding can be found by fixating γ at its mean :

$$\hat{\underline{d}}(\mathcal{E})^{a} = \underset{\underline{d}}{\max}^{-1}\left(\langle v\lfloor \underline{d},\mathcal{E}\rangle - \frac{1}{2}\langle\gamma|\mathcal{E}\rangle\, \overset{v}{\sim}v\lfloor \underline{d},\mathcal{E}\rangle\right) \qquad (3.3.13)$$

$$\hat{\underline{d}}(\gamma,\mathcal{E})^{a} = \underset{\underline{d}}{\max}^{-1}\left(\langle v\lfloor \underline{d},\mathcal{E}\rangle - \frac{1}{2}\,\gamma\, \overset{v}{\sim}v\lfloor \underline{d},\mathcal{E}\rangle\right) \qquad (3.3.14)$$

The prior solution (3.3.13) follows from Section 3.2 by substituting $\langle\gamma|\mathcal{E}\rangle$ for γ :

$$\underline{\hat{d}}(\mathcal{E})^{a} = \underline{z}\,\langle\gamma|\mathcal{E}\rangle \tag{3.3.15}$$

$$\langle v|\underline{\hat{d}}(\mathcal{E})^{a},\mathcal{E}\rangle^{a} = a + \frac{1}{2}\langle\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle + \frac{1}{2}\underline{z}'\underline{H}\,\underline{z}\,\langle\gamma|\mathcal{E}\rangle^{2} \tag{3.3.16}$$

where for notational convenience we define

$$\underline{z} = (\underline{H} - \gamma\,\underline{G}'\underline{C}\,\underline{G})^{-1}\,\underline{G}'(\underline{C}\,\underline{b} + \frac{1}{2}\langle\underline{s}\,\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle) \tag{3.3.17}$$

The solution when γ is known is

$$\underline{\hat{d}}(\gamma,\mathcal{E})^{a} = \underline{z}\,\gamma \tag{3.3.18}$$

$$\langle v|\underline{\hat{d}}(\gamma,\mathcal{E})^{a},\gamma,\mathcal{E}\rangle = a + \frac{1}{2}\langle\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle + \frac{1}{2}\underline{z}'\underline{H}\,\underline{z}\,\gamma^{2}\,. \tag{3.3.19}$$

Taking the expectation of (3.3.19) we have

$$\langle\langle v|\underline{\hat{d}}(\gamma,\mathcal{E})^{a},\gamma,\mathcal{E}\rangle|\mathcal{E}\rangle = a + \frac{1}{2}\langle\underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle + \frac{1}{2}\underline{z}'\underline{H}\,\underline{z}\,\langle\gamma^{2}|\mathcal{E}\rangle. \tag{3.3.20}$$

Substituting (3.3.20) and (3.3.16) into (3.3.12) we have the result:

$$\langle v_{\gamma}|\mathcal{E}\rangle^{am} = -\frac{1}{2}\underline{z}'\underline{H}\,\underline{z}\,{}^{v}\langle\gamma|\mathcal{E}\rangle \approx -\frac{1}{2}\underline{b}'\underline{C}\,\underline{G}(\underline{H} - \gamma\,\underline{G}'\underline{C}\,\underline{G})^{-1}\underline{G}'\,\underline{C}\,\underline{b}\,{}^{v}\langle\gamma|\mathcal{E}\rangle \tag{3.3.21}$$

where we have dropped the terms involving third covariances from (3.3.21).

<u>Discussion of the Value of Additional Risk Preference</u>

We can rederive (3.3.21) using the theorem of Chapter 2 and the following analogy:

$$\langle v|\underline{d},\mathcal{E}\rangle^{an} = {}^{\sim}\langle v|\underline{d},\mathcal{E}\rangle \tag{3.3.22}$$

$$\underline{s}^{an} = \gamma - \langle\gamma|\mathcal{E}\rangle \tag{3.3.23}$$

$$\underline{d}^{an} = \underline{d} \tag{3.3.24}$$

where the superscript an denotes analogy. We treat the certain equivalent as the mean or objective function and γ as the state variable, leaving $\underline{d}$ as the decision vector. Differentiating the objective function and evaluating at $\underline{d} = \underline{0}$ and $\gamma = \langle\gamma|\mathcal{E}\rangle$, we have

$$\underline{H}^{an} = \nabla^2 \tilde{v}\lfloor \underline{d},\mathcal{E}\rangle = \underline{H} - \langle\gamma|\mathcal{E}\rangle\, \underline{G}'\underline{C}\,\underline{G} \tag{3.3.25}$$

$$\underline{G}^{an} = \nabla\left(\frac{\partial\, \tilde{v}\lfloor \underline{d},\mathcal{E}\rangle}{\partial\gamma}\right) = \underline{b}'\underline{C} + \langle \underline{s}'\underline{E}\,\underline{s}\,\underline{s}'|\mathcal{E}\rangle\,. \tag{3.3.26}$$

Notice that the covariance matrix $\underline{C}$ has been absorbed into the coefficients. Finally the prior variance of the posterior mean of the state variable is

$$^{v}\langle\langle s|D,\mathcal{E}\rangle^{an}|\mathcal{E}\rangle = {}^{v}\langle\langle\gamma|\gamma,\mathcal{E}\rangle|\mathcal{E}\rangle = {}^{v}\langle\gamma|\mathcal{E}\rangle\,. \tag{3.3.27}$$

Substituting (3.3.25), (3.3.26) and (3.3.27) into (2.3.9) and dropping the terms involving third covariances, we have

$$\begin{aligned}\langle v_{\gamma}|\mathcal{E}\rangle^{an} &\approx -\frac{1}{2}\,\underline{G}^{an}\,\underline{H}^{an^{-1}}\,\underline{G}^{an'}\;{}^{v}\langle\gamma|\mathcal{E}\rangle \\ &= -\frac{1}{2}\,\underline{b}'\underline{C}\,(\underline{H} - \langle\gamma|\mathcal{E}\rangle\,\underline{G}'\underline{C}\,\underline{G})^{-1}\,\underline{C}\,\underline{b}\;{}^{v}\langle\gamma|\mathcal{E}\rangle\,.\end{aligned} \tag{3.3.28}$$

which is the same as (3.3.21).

The conclusion is that by treating γ as if it were a state variable we can find the value of additional encoding. The parameters required other than $\langle\gamma|\mathcal{E}\rangle$ and ${}^{v}\langle\gamma|\mathcal{E}\rangle$ are the same ones needed to compute the value of clairvoyance on the state variables.

CHAPTER 4

THE DESIGN OF MONTE CARLO SAMPLING

4.0 Introduction

This chapter is logically separate from the others. To understand the applications in Chapter 5, the reader must comprehend this development as well as the results in Chapters 2 and 3. However, in this chapter only the results in Section 4.1 are important to the reader's understanding of Chapter 5. The derivation of Section 4.2 and the discussion of Section 4.3 involve new concepts and notation which might be a burden to the casual reader.

4.1 The Expected Loss from Incomplete Sampling

In this section we discuss how sampling can be used to approximate the optimum decision for a problem with a single decision variable. We define the resulting loss as the difference between the exact expected value and the approximate one. We state and discuss the result for a quadratic value function in this section, leaving the derivation until Section 4.2.

The deterministic model for this section has one decision variable and many state variables :

$$v(\underline{s},d) = a + \underline{b}'\underline{s} + \frac{1}{2}\,\underline{s}'\underline{E}\,\underline{s} + \underline{s}'\underline{g}\,d + \frac{1}{2}\,h\,d^2 \qquad (4.1.1)$$

where we have modified the $\underline{G}$ and $\underline{H}$ of previous chapters to $\underline{g}$ and h respectively. This reflects the change in dimensionality. The object of the sampling program is to maximize the expected value of v :

$$\max_{d} \langle v|d,\mathcal{E}\rangle \tag{4.1.2}$$

For the quadratic problem we can solve (4.1.2) exactly using the results of Chapter 2. However, for more complex problems the exact solution may be impossible to find, and we might use an approximate solution based on sampling. The general approach is to discretize the decision variable and then to find the approximate expected value $\langle v|d_j,\mathcal{E}\rangle^a$ at each discrete decision setting d_j. We can generate the i^{th} sample at the j^{th} decision setting by choosing a random sample from the probability density function of the state vector $\{\underline{s}|\mathcal{E}\}$ and calculating the associated value of v :

$${}_iv_j = v({}_i\underline{s},d_j) \tag{4.1.3}$$

Sampling in this manner, we can solve for $\hat{d}(\mathcal{E})$ without ever calculating the probability density function $\{v|d,\mathcal{E}\}$

Figure 4.1 illustrates the terminology we need to state the sampling problem more precisely. The n_d decision settings are equally spaced over a predetermined range 2Δ. The range is centered at the deterministic optimum, which is zero in accord with previous chapters:

$$\hat{d}_o = 0 \tag{4.1.4}$$

To simplify the notation and derivation of Section 4.2, we have assumed that n_d is odd and defined the new parameter L, where

$$n_d = 2L + 1 . \tag{4.1.5}$$

At each decision setting d_j, n_s random samples are taken from the probability density function $\{v|d_j,\mathcal{E}\}$. Then a quadratic least

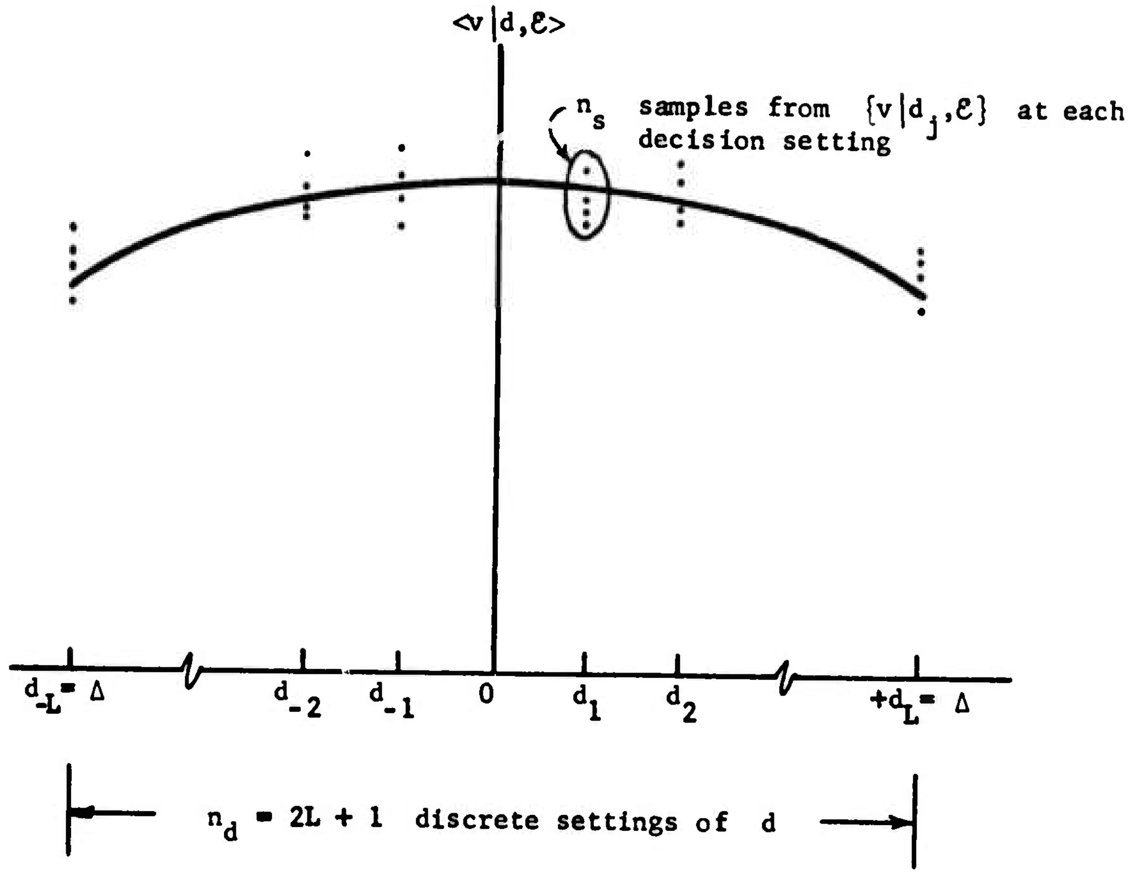


Figure 4.1 Terminology for the sampling problem

squares curve is fit through all of the data points. The maximum of this curve $\hat{d}(\mathcal{E})^a$ approximates the optimum decision $\hat{d}(\mathcal{E})$.

The main result of this section is that for large n_d the expected loss $<\ell|N,\mathcal{E}>$ from using $\hat{d}(\mathcal{E})^a$ rather than $\hat{d}(\mathcal{E})$ is

$$<\ell|N,\mathcal{E}> = \frac{3}{4\,h\,\Delta^2}\;\frac{\overset{v}{<}v|\hat{d}(\mathcal{E}),\mathcal{E}>}{N} \tag{4.1.6}$$

where

$$N = n_s n_d \; . \tag{4.1.7}$$

The loss is proportional to the variance at $\hat{d}(\mathcal{E})$ divided by the total number of samples. This is the result that we need for Chapter 5.

Since the loss in (4.1.6) depends on the product of n_s and n_d , we would not expect it to matter whether we discretize d finely, taking only a few samples per setting, or whether we discretize d coarsely, taking many samples per setting. However, this conclusion is only valid for large n_d . In other words, very fine discretization is approximately equivalent to fine discretization. In Section 4.3 we show that fine discretization always results in a smaller expected loss than coarse discretization.

Finally, let us recognize that (4.1.6) is based on a simple sampling program. More complicated procedures would use higher order curve fits and would combine prior data with sample data. As long as the range Δ is not too large and the sample size N is not too small, those sophisticated techniques should still exhibit the insensitivity of our simple approach.

4.2 Proof that the Expected Loss Is Inversely Proportional to the Total Number of Sample Points

To derive expression (4.1.6), we need to introduce some additional concepts. Least squares curve fitting is most conveniently analyzed in terms of orthogonal polynomials. Any function $f(d)$ that can be expanded in a Taylor series can also be expanded in terms of orthogonal polynomials :

$$f(d) = \sum_{k=0}^{m} b_k \varphi_k(d) \tag{4.2.1}$$

The b_k's are the coefficients, and the φ_k's are the polynomials. The number of terms m is the order of the fit.

Ralston [8] shows that the first three orthogonal polynomials are:

$$\varphi_0 = 1 \tag{4.2.2}$$

$$\varphi_1 = d/\Delta \tag{4.2.3}$$

$$\varphi_2 = -\frac{L+1}{2L-1} + \frac{3d^2L}{(2L-1)\Delta^2} \tag{4.2.4}$$

where Δ and L are defined in Section 4.1. If we sample $f(d)$ at the points d_j , where j varies over the integers from $-L$ to $+L$, the sample coefficients are defined as

$$b_k^a = \frac{\sum_{j=-L}^{+L} f(d_j)\ \varphi_k(d_j)}{\sum_{j=-L}^{L} \varphi_k^2(d_j)} . \tag{4.2.5}$$

Ralston shows that the sample coefficients are unbiased; that is, if we know the exact coefficients in (4.2.1), then

$$\langle b_k^a|\mathcal{E}\rangle = b_k \ . \tag{4.2.6}$$

Ralston also computes the variances of the coefficients for the case where the variance of $f(d)$ is independent of d :

$${}^{v}\langle f|d,\mathcal{E}\rangle = {}^{v}\langle f|\mathcal{E}\rangle = \sigma^2 \tag{4.2.7}$$

The first three variances are :

$${}^{v}b_0^a = \frac{\sigma^2}{2L + 1} \tag{4.2.8}$$

$${}^{v}b_1^a = \frac{3L\ \sigma^2}{(L + 1)(2L + 1)} \tag{4.2.9}$$

$${}^{v}b_2^a = \frac{10L(2L - 1)\sigma^2}{(2L + 3)(2L + 2)(2L + 1)} \tag{4.2.10}$$

By a derivation similar to the one for the variances of the b's (see Ralston [8, p. 247]), it is straightforward to show that the covariances are zero :

$${}^{cov}\langle b_k b_\ell|\mathcal{E}\rangle = 0 \ , \quad k \neq \ell \tag{4.2.11}$$

Specializing Ralston's results to our problem, $\langle v|d,\mathcal{E}\rangle$ plays the role of $f(d)$. Taking the expectation of (4.1.1), we have

$$\langle v|d,\mathcal{E}\rangle = a + \frac{1}{2}\langle \underline{s}'\underline{E}\ \underline{s}|\mathcal{E}\rangle + \frac{1}{2}\, h\, d^2 \ . \tag{4.2.12}$$

Fitting this curve using the orthogonal polynomials (4.2.2), (4.2.3) and (4.2.4), the exact coefficients are :

$$b_0 = a + \frac{1}{2}\langle \underline{s}'\underline{E}\ \underline{s}|\mathcal{E}\rangle + \frac{(L + 1)h\ \Delta^2}{6L} \tag{4.2.13}$$

$$b_1 = 0 \tag{4.2.14}$$

$$b_2 = \frac{(2L - 1)h\ \Delta^2}{6L} \tag{4.2.15}$$

By direct calculation or by reference to Chapter 2, the exact solution is

$$\max_d \langle v|d,\mathcal{E}\rangle = \langle v|\hat{d}(\mathcal{E}),\mathcal{E}\rangle = a + \frac{1}{2} \langle \underline{s}'\underline{E}\ \underline{s}|\mathcal{E}\rangle\ . \tag{4.2.16}$$

Suppose that instead of calculating the coefficients exactly, we approximate them by the sampling procedure suggested in Section 4.1. Define ${}_iv_j$ as the i^{th} sample at the j^{th} decision point. Then we can compute the sample expectation at each decision setting as

$$\langle v|d_j,\mathcal{E}\rangle^a = \sum_{i=1}^{n_s} {}_iv_j/n_s\ , \tag{4.2.17}$$

where n_s is the number of samples per decision setting. If we substitute (4.2.17) into (4.2.5) as $f(d_j)$ and take the expectation, we find after some algebraic manipulations,

$$\langle b_k^a|\mathcal{E}\rangle = b_k\ , \quad k = 0,1,2\ , \tag{4.2.18}$$

where the b_k's are defined by (4.2.13), (4.2.14) and (4.2.15). In other words $\langle v|d_j,\mathcal{E}\rangle^a$ as defined by (4.2.17) is an unbiased estimator of $\langle v|d_j,\mathcal{E}\rangle$, and therefore Ralston's results apply.

The expected loss from using the approximate solution is

$$\langle \ell|N,\mathcal{E}\rangle = \langle v|\hat{d}(\mathcal{E}),\mathcal{E}\rangle - \left\langle \langle v|\hat{d}(\mathcal{E})^a,\mathcal{E}\rangle^a|\mathcal{E}\right\rangle\ . \tag{4.2.19}$$

The first term on the right of (4.2.19) is the exact solution from (4.2.16). The second term is the expected solution based on sampling. To find an expression for it, we define the approximate expected value

based on the sample data as

$$\langle v|d,\mathcal{E}\rangle^{a} = \sum_{k=0}^{2} b_k^{a}\,\varphi_k(d) \; . \qquad (4.2.20)$$

Maximizing (4.2.20), using (4.2.2), (4.2.3) and (4.2.4) for the polynomials, we find the approximate optimal decision $\hat{d}(\mathcal{E})^{a}$:

$$\hat{d}(\mathcal{E})^{a} = -\frac{2L-1}{6L}\,\frac{b_1^{a}}{b_2^{a}} \qquad (4.2.21)$$

Substituting (4.2.21) into (4.2.20), and simplifying, we have

$$\langle v|\hat{d}(\mathcal{E})^{a},\mathcal{E}\rangle^{a} = b_0^{a} - \frac{L+1}{2L-1}\,b_2^{a} - \frac{2L-1}{12L}\,\frac{(b_1^{a})^2}{b_2^{a}} \; . \qquad (4.2.22)$$

To take the expectation of (4.2.22), we expand the reciprocal of b_2^{a} in a Taylor series about the mean $\bar{b}_2{}^{a}$:

$$\frac{1}{b_2^{a}} \cong \frac{1}{\bar{b}_2{}^{a}}\left(1 + \frac{1}{(\bar{b}_2{}^{a})^2}\,(b_2^{a} - \bar{b}_2{}^{a})^2\right) \qquad (4.2.23)$$

Taking the expectation of (4.2.22) and using (4.2.23), we have

$$\langle\langle v|\hat{d}(\mathcal{E})^{a},\mathcal{E}\rangle^{a}|\mathcal{E}\rangle \cong \bar{b}_0{}^{a} + \frac{L+1}{2L-1}\,\bar{b}_2{}^{a} - \frac{2L-1}{12L}\,\frac{\overset{v_a}{b_1}}{\bar{b}_2{}^{a}}\left(1 - \frac{\overset{v_a}{b_2}}{(\bar{b}_2{}^{a})^2}\right) . \qquad (4.2.24)$$

Recalling that the means are given by (4.2.13) through (4 2.15) and using (4.2.8) through (4.2.10) for the variances, (4.2.24) becomes

$$\langle\langle v|\hat{d}(\mathcal{E})^{a},\mathcal{E}\rangle|\mathcal{E}\rangle = a + \frac{1}{2}\langle \underline{s}'\underline{E}\;\underline{s}|\mathcal{E}\rangle - \frac{3L\,\sigma^2}{(2L+2)(2L+1)h\,\Delta^2} \qquad (4.2.25)$$

$$+ \left(1 - \frac{720L^4\sigma^2}{(2L+3)(2L+2)(2L+1)(2L)(2L-1)h^2\,\Delta^4}\right).$$

Expression (4.1.6) follows directly from (4.2.25) by letting L and hence n_d become large.

4.3 Discussion of the Expected Loss from Rough Quantization

Expression (4.2.22) provides a more sophisticated basis than (4.1.4) for discussing rough quantization. We define the sharpness of the maximum at $\hat{d}(\mathcal{E})$ as

$$\lambda = \frac{\Delta v}{\sigma} \tag{4.3.1}$$

where

$$\Delta v = \langle v|\hat{d}(\mathcal{E}),\mathcal{E}\rangle - \langle v|d=\Delta,\mathcal{E}\rangle \tag{4.3.2}$$

$$\sigma^2 = {}^{v}\langle v|\hat{d}(\mathcal{E}),\mathcal{E}\rangle\ . \tag{4.3.3}$$

Using (4.2.10) for the quadratic value function, (4.3.1) becomes

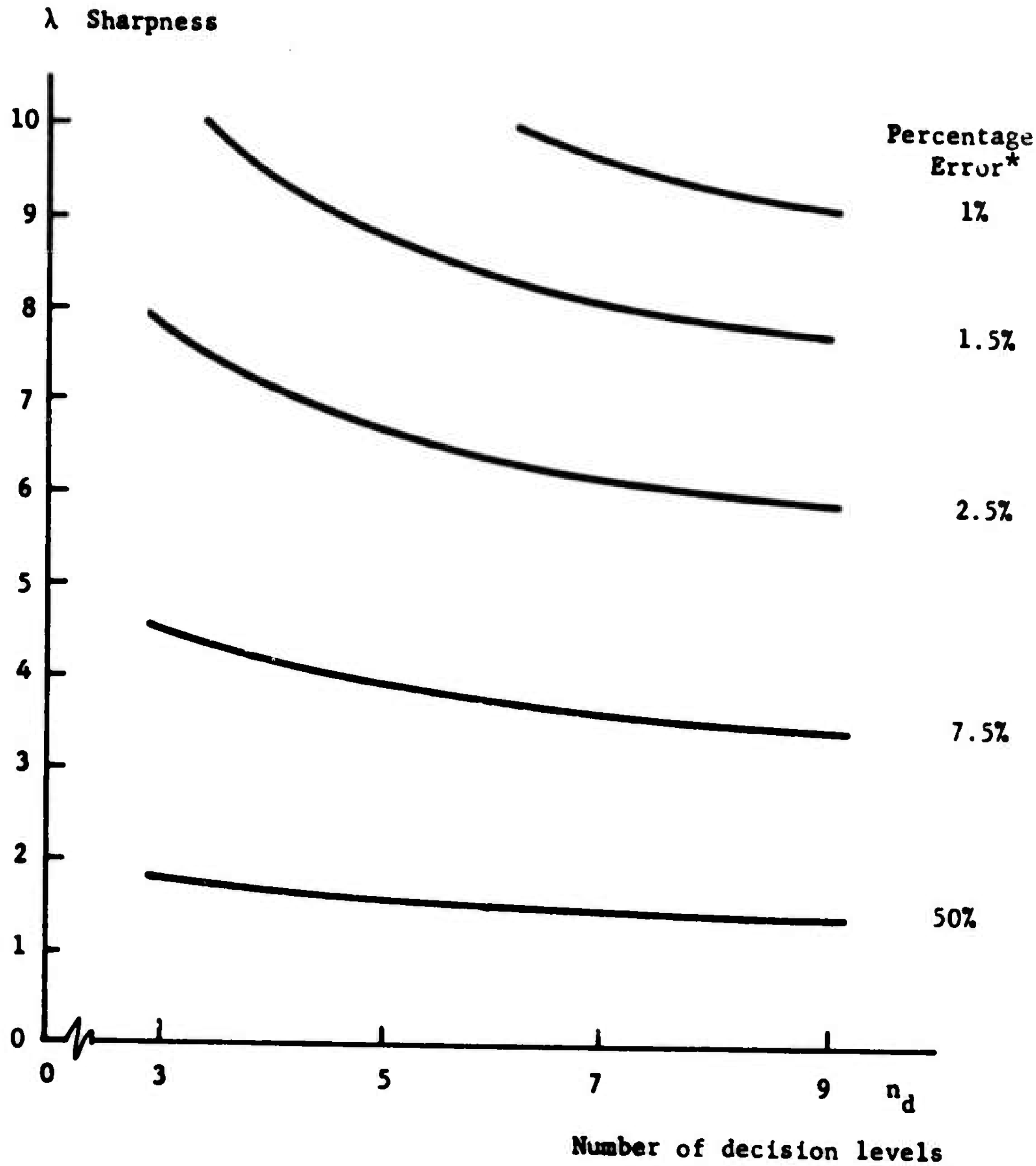
$$\lambda = -\frac{1}{2}\frac{h\ \Delta^2}{\sigma}\ . \tag{4.3.4}$$

Using (4.3.4), the final term in (4.2.25) may be written

$$\frac{180\ L^4\ \lambda^2}{(2L+3)(2L+2)(2L+1)(2L)(2L-1)}\ . \tag{4.3.5}$$

For large N , this term represents the error that is introduced by using (2L + 1) quantization levels instead of N levels.

The term in (4.3.5) is plotted in Fig. 4.2. We see that for sharp maxima rough quantization introduces little error. For maxima that are not sharp, the error is larger but insensitive to quantization. We conclude that although fine quantization is always better than rough quantization, for practical purposes the sampling error depends only on the total number of samples and not on the coarseness of the quantization.



*Relative to one sample per level for large N

Figure 4.2 Quantization error for a decision variable as a function of the sharpness of the maximum and the number of quantization levels.

CHAPTER 5

APPLICATIONS

5.0 Introduction

This chapter is the focal point of the thesis. In Sections 5.1 through 5.4 we apply the results of Chapters 2, 3 and 4 to develop a systematic design framework. It applies to problems with continuous decision variables. The design is optimal in the sense that our goal is to make the marginal benefit of additional analysis equal to marginal cost. In section 5.5 we discuss how the framework could be modified to apply to budget constrained design and to problems with discrete decision variables.

5.1 Preliminary Analysis

In Sections 5.1 through 5.4 we consider the problem introduced in Section 2.3. The deterministic model $v(\underline{s},\underline{d})$ can be approximated by a second order Taylor series about the mean of the state variables and the prior optimum decision. The model $v(\underline{s},\underline{d})$ may be very complex. A single evaluation of $v(\underline{s},\underline{d})$ on a computer may cost many dollars.

To perform an exhaustive probabilistic analysis to find the exact optimum decision $\hat{\underline{d}}(\mathcal{E})$ would be prohibitively expensive. Our objective is to use preliminary data to identify cost effective additional analyses. As input to our framework we require roughly encoded parameters and deterministic sensitivity data. As output we recommend the level of encoding for state variables, the proper detail of the treatment of risk preference, and the amount of computation. We use

approximations to keep the cost and input requirements of our analysis to a minimum.

The Three Steps in the Preliminary Phase

Figure 5.1 summarizes the preliminary phase. Deterministic sensitivity analysis yields the deterministic optimum $\underline{\hat{d}}_o$ and the first and second partial derivatives of the state and decision variables at the operating point $(\underline{\bar{s}},\underline{\hat{d}}_o)$. For a general discussion of sensitivity analysis, see Howard [1]. For a specific discussion of the conversion of sensitivity data to approximate partial derivatives, see Howard [2, Appendix A].

Once we have the sensitivity data we can normalize the state and decision variables so that they are zero at the mean and deterministic optimum respectively :

$$\underline{\bar{s}} = \underline{0} \tag{5.1.1}$$

$$\underline{\hat{d}}_o = \underline{0} \tag{5.1.2}$$

This step is not essential, but it simplifies notation and data handling. Using the sensitivity data and the normalized variables, we can fit a quadratic Taylor series model to $v(\underline{s},\underline{d})$ at $(\underline{s},\underline{d}) = (\underline{0},\underline{0})$:

$$v(\underline{s},\underline{d}) = a + \underline{b}'\underline{s} + \frac{1}{2}\,\underline{s}'\underline{E}\,\underline{s} + \underline{s}'\underline{G}\,\underline{d} + \frac{1}{2}\,\underline{d}'\underline{H}\,\underline{d} \tag{5.1.3}$$

The final step is to encode the matric of covariances. Howard [2, p. 511] suggests an encoding technique. In many problems the state variables will be independent, reducing the task to encoding the variances of the state variables. In this case we may directly encode rough estimates of the variances, or we may estimate them based on the

START

DETERMINISTIC SENSITIVITY ANALYSIS

Perform deterministic sensitivity analysis to find $\hat{\underline{d}}_o$ the deterministic optimum and the first and second order partial derivatives at $\underline{d} = \hat{\underline{d}}_o$ and $\underline{s} = \overline{\underline{s}}$

Ref: Howard [2, Appendix A]

NORMALIZE AND FIT

Normalize the state and decision vectors such that

$$\overline{\underline{s}} = 0$$

$$\hat{\underline{d}}_o = 0$$

and fit

$$v(\underline{s},\underline{d}) = a+\underline{b}'\underline{s} + \frac{1}{2}\,\underline{s}'\underline{E}\,\underline{s} + \underline{s}'\underline{G}\,\underline{d} + \frac{1}{2}\,\underline{d}'\underline{H}\,\underline{d}$$

Ref: Section 2.3

ENCODE COVARIANCES

Encode the covariance matrix

Ref: Howard [2, p. 511]

or

Encode A and default to

$$\mathrm{cov}_{ij} = 0, \quad i \neq j$$

$$\mathrm{cov}_{ij} = \overset{v}{s}_i = \ell_i^2/A, \quad i = j$$

Ref: Section 5.2

GO TO ENCODING DECISION

Figure 5.1 Preliminary analysis

ranges used for the sensitivity analysis. Conversion from a range to a variance involves the parameter A which is discussed in Section 5.2.

Illustrative Example, The Entrepreneur's Problem

To illustrate the methodology of this chapter we shall apply each new procedure to the Entrepreneur's Problem, which was introduced in Chapter 2. Expressing the profit π in millions of dollars and denoting the cost Δc as s_1 and the quantity Δq as s_2, the coefficients of (5.1.3) are :

$$a = 198 \tag{5.1.4}$$

$$\underline{b}' = [-1 \quad 17.58] \tag{5.1.5}$$

$$\underline{E} = \begin{bmatrix} 0 & 0 \\ 0 & 0.0497 \end{bmatrix} \tag{5.1.6}$$

$$\underline{G}' = [0 \quad 0.835] \tag{5.1.7}$$

$$\underline{H} = [\ -3.67\ ] \tag{5.1.8}$$

The covariance matrix and the risk aversion coefficient are :

$$\underline{C} = \begin{bmatrix} 10{,}000 & 0 \\ 0 & 100 \end{bmatrix} \tag{5.1.9}$$

$$\gamma = 0.004 \tag{5.1.10}$$

The Validity of the Approximations

Once we complete the preliminary analysis, we check to see if the approximations developed in Chapters 2 and 3 are applicable. Figure 5.2 illustrates the formal checks. In Section 3.2 we developed expressions

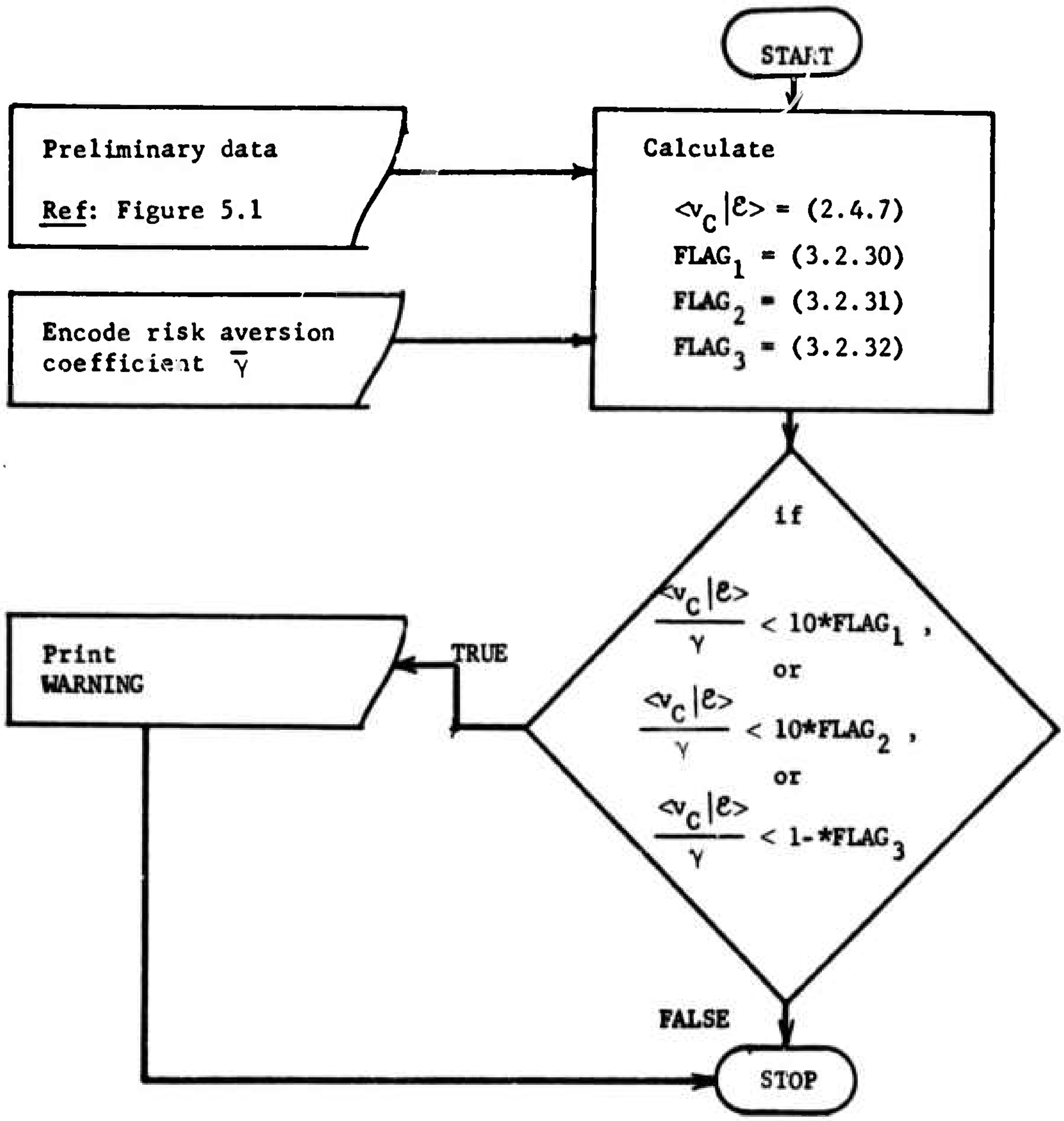


Figure 5.2 Checks to see if approximations are valid

for three quantities that should be small relative to the value of information of the state variables. If these conditions are not met, Fig. 5.2 indicates that the user should be warned. In this situation, careful modification of the framework, perhaps using the assumption of Section 3.2 that state variables are normal and independent, may salvage the analysis. To make the framework general enough to handle violations of the conditions would make it too complex to be of practical use.

An informal check should be applied at this point. The expressions of Chapters 2 and 3 assume that the probability density functions are roughly centrally symmetric. If the user feels that any third covariances are large, he should proceed with caution. The expressions of Chapter 3, particularly (3.2.22), should help the user to assess the seriousness of an asymmetry.

Application of the Checks to the Entrepreneur's Problem

Applied to the Entrepreneur's Problem the three checks of Fig. 5.2 are :

$$\frac{\langle v_c|\mathcal{E}\rangle}{\gamma} = 26 * \mathrm{FLAG}_1 \tag{5.1.11}$$

$$\frac{\langle v_c|\mathcal{E}\rangle}{\gamma} = 50 * \mathrm{FLAG}_2 \tag{5.1.12}$$

$$\frac{\langle v_c|\mathcal{E}\rangle}{\gamma} = 2 * \mathrm{FLAG}_3 \tag{5.1.13}$$

As we shall see in Section 5.3 the low value of 2 in the third check is an early warning that the Entrepreneur's Problem is highly risk sensitive.

To check for the impact of asymmetry, suppose that the probability

distribution on Δq is lognormal with the same mean and variance as before. The lower bound of the probability density function is -68 , corresponding to $q = 0$. From (3.2.22) third covariances are unimportant if

$$\langle \underline{s}'\underline{E}\ \underline{s}\ \underline{s}' | \mathcal{E}\rangle \ll \underline{b}'\underline{C} \ . \tag{5.1.14}$$

Computing these quantities the inequality is

$$[0 \quad 26] \ll [0 \quad 1758] \ . \tag{5.1.15}$$

The assumed asymmetry produces negligible changes in the results.

5.2 Encoding State Variables

The first analytical option is whether to gather additional data about the state variables. In theory the data about the state variables could come from a variety of sources; from a simulation model, from an experiment, or from an expert. As long as the prior variance of the posterior mean $\langle\langle v|D,\mathcal{E}\rangle|\mathcal{E}\rangle$ can be assessed, the evaluation scheme of Fig. 5.3 applies. In practice the most likely application is where the data comes from an encoding interview with the decision maker or his designated experts. At the end of this section we discuss how the input data might be generated for this case.

The iterative encoding procedure of Fig. 5.3 applies when the state variables are independent and the cost of encoding each variable is a constant K_e . If the encoding costs were a function of n the number of questions in an encoding interview, then the preposterior variance would have to be specified as a function of n . This would require experimental research on how the mean of a decision maker's

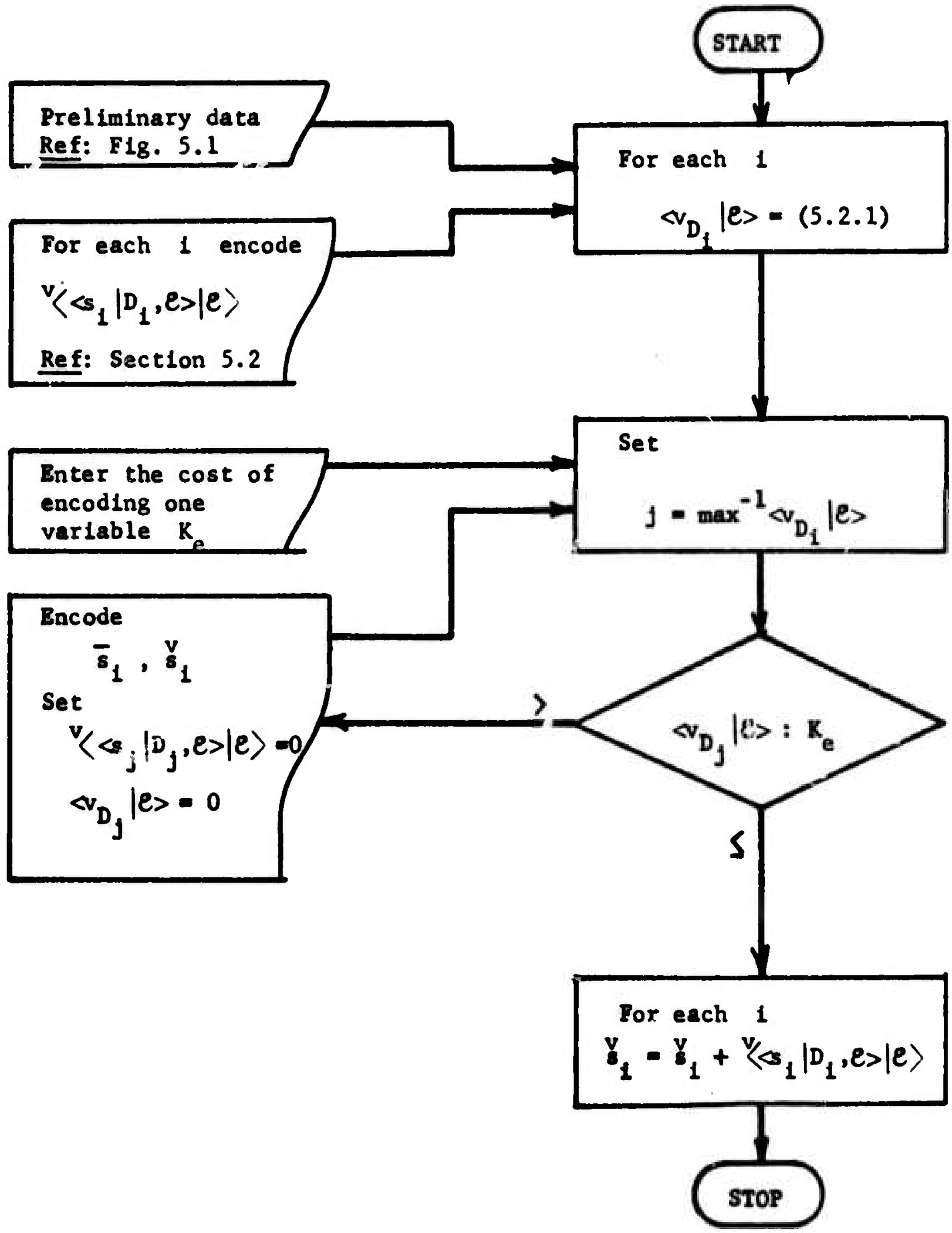


Figure 5.3 Iterative encoding of state variables

distribution varies during an interview. We leave this topic for further research. For our analysis we assume that we either use the preliminary estimate or encode a distribution which the decision maker accepts as representing his state of information.

The calculation for the value of the data is based on the ranking scheme from Chapter 2. Assuming that encoding a probability distribution on one variable gives no information about other variables, the value of data on the i^{th} variable $\langle v_{D_i}|\mathcal{E}\rangle$ is given by (2.4.8) specialized to the case where only the i^{th} variance is non-zero:

$$\langle v_{D_i}|\mathcal{E}\rangle = -\frac{1}{2}\,g_i' H^{-1} g_i \, {}^{v}\langle\langle s_i|D_i,\mathcal{E}\rangle|\mathcal{E}\rangle \qquad (5.2.1)$$

The heart of the encoding procedure is the iterative loop in the middle of Fig. 5.3. We sort the values of encoding to form a list with the largest value $\langle v_{D_j}|\mathcal{E}\rangle$ at the top. If the value of encoding exceeds the cost we encode the j^{th} variable completely, yielding a new mean and variance. The variance ${}^{v}\langle\langle s_j|D_j,\mathcal{E}\rangle|\mathcal{E}\rangle$ is zero after the complete encoding, dropping the j^{th} variable to the bottom of the list. We continue until the maximum value of encoding is less than the cost.

The iterative procedure is based on the assumption that the joint value of clairvoyance on two state variables is equal to the sum of the individual values of clairvoyance. The assumption is good if the joint value does not greatly exceed the marginal values.

At the end of Fig. 5.3 we set the variance of any unencoded variable to the sum of the point estimate of the variance and the prior variance of the posterior mean. These quantities are discussed in Section 2.4.

Encoding Variances

The key to the encoding decision is estimating ${}^{v}\langle\langle s_i|D_i,\mathcal{E}\rangle|\mathcal{E}\rangle$ and $\langle{}^{v}\langle s_i|D_i,\mathcal{E}\rangle|\mathcal{E}\rangle$. Since both are measures of the dispersion of the distribution, they are related to ℓ_i , the difference between the high and low estimates for s_i .

If $\{s_i|\mathcal{E}\}$ is uniformly distributed between the high and low estimates, the variance is

$$ {}^{v}\langle s_i|\mathcal{E}\rangle = \frac{\ell_i^2}{12} \ . \tag{5.2.2} $$

If $\{s_i|\mathcal{E}\}$ is normally distributed with the high and low values each three standard deviations from the mean, the variance is

$$ {}^{v}\langle s_i|\mathcal{E}\rangle = \frac{\ell_i^2}{36} \ . \tag{5.2.3} $$

A reasonable model of the relationship between the variance and ℓ_i is

$$ {}^{v}\langle s_i|\mathcal{E}\rangle = \frac{\ell_i^2}{A} \ . \tag{5.2.4} $$

Assuming that ${}^{v}\langle\langle s_i|D_i,\mathcal{E}\rangle|\mathcal{E}\rangle$ and $\langle{}^{v}\langle s_i|D,\mathcal{E}\rangle|\mathcal{E}\rangle$ can be expressed as ℓ_i^2 divided by A_{VE} and A_{EV} respectively and using (2.4.19), the variance is

$$ {}^{v}\langle s_i|\mathcal{E}\rangle = \frac{\ell_i^2}{A} \tag{5.2.5} $$

where

$$ \frac{1}{A} = \frac{1}{A_{VE}} + \frac{1}{A_{EV}} \ . \tag{5.2.6} $$

Of course, assignment of A_{VE} and A_{EV} for any given problem depends

on how the decision maker or his designated analyst interpret the terms "high value" and "low value." Hopefully, during future decision analyses, data can be gathered on the relationship between A_{VE} and A_{EV} .

When we assume that A_{EV} and A_{VE} are the same for each variable, our encoding procedure ranks the variables by their approximate values of clairvoyance. Even in this simple case, our analysis gives us insight. A common misconception is to assume that the importance of a state variable is measured by the first partial derivative of the value function. In fact, the first partial derivative has nothing to do with the value of data for the risk indifferent problem with a quadratic value function.

Encoding in the Entrepreneur's Problem

Rather than specifying the prior variance of the posterior mean directly, we parameterize the solution on the ratio r :

$$r = \frac{{}^{v}\big\langle\langle v|D,\mathcal{E}\rangle|\mathcal{E}\big\rangle}{{}^{v}\langle v|\mathcal{E}\rangle} = \frac{A_{EV}}{A_{VE}} \tag{5.2.7}$$

The values of encoding for the state variables in the Entrepreneur's Problem are

$$\langle v_{D_1}|\mathcal{E}\rangle = 0 \tag{5.2.8}$$

$$\langle v_{D_2}|\mathcal{E}\rangle = 9.5\, r\ . \tag{5.2.9}$$

The value of encoding the cost s_1 is zero because the partial derivative of profit with respect to cost and price is zero. Since the Entrepreneur's Problem does not include the option to stop, the costs

will be incurred regardless of the pricing decision. There is no value in learning the exact value of the sunk costs.

The cost of encoding K_e should be approximately a thousand dollars or 10^{-3} million dollars. Since we expect r to be in the range of 0.1 to 1.0 for a typical problem, the decision is clearly to encode s_2 and not to encode s_1 .

5.3 <u>The Choice of Risk Attitude</u>

Risk preference follows the encoding of state variables both in the chronology of decision analysis and in the complexity of computation. In this section we choose the appropriate risk attitude for the probability phase based on preliminary estimates of the risk aversion coefficient.

The options for the probabilistic phase are:

(i) <u>Linear utility</u> where γ is zero, the decision maker is <u>risk-indifferent</u>.

(ii) <u>Exponential utility</u> where γ is fixed, the decision maker has <u>constant risk preference</u>.

(iii) <u>Complete utility</u> where γ is a function of wealth, the decision maker has <u>wealth-sensitive risk preference</u>.

<u>The Flow Chart</u>

Figure 5.4 summarizes the evaluation of the alternatives. The preliminary attitude is summarized by the risk preference coefficient $\bar{\gamma}$. The potential risk preference coefficients after a thorough encoding of the decision maker's risk attitude and an accurate computation of the profit lottery are summarized by $\check{\gamma}$.

The costs K_{12} and K_{23} include education, assessment, and computation relative to option (ii). The negative of K_{12} is the savings

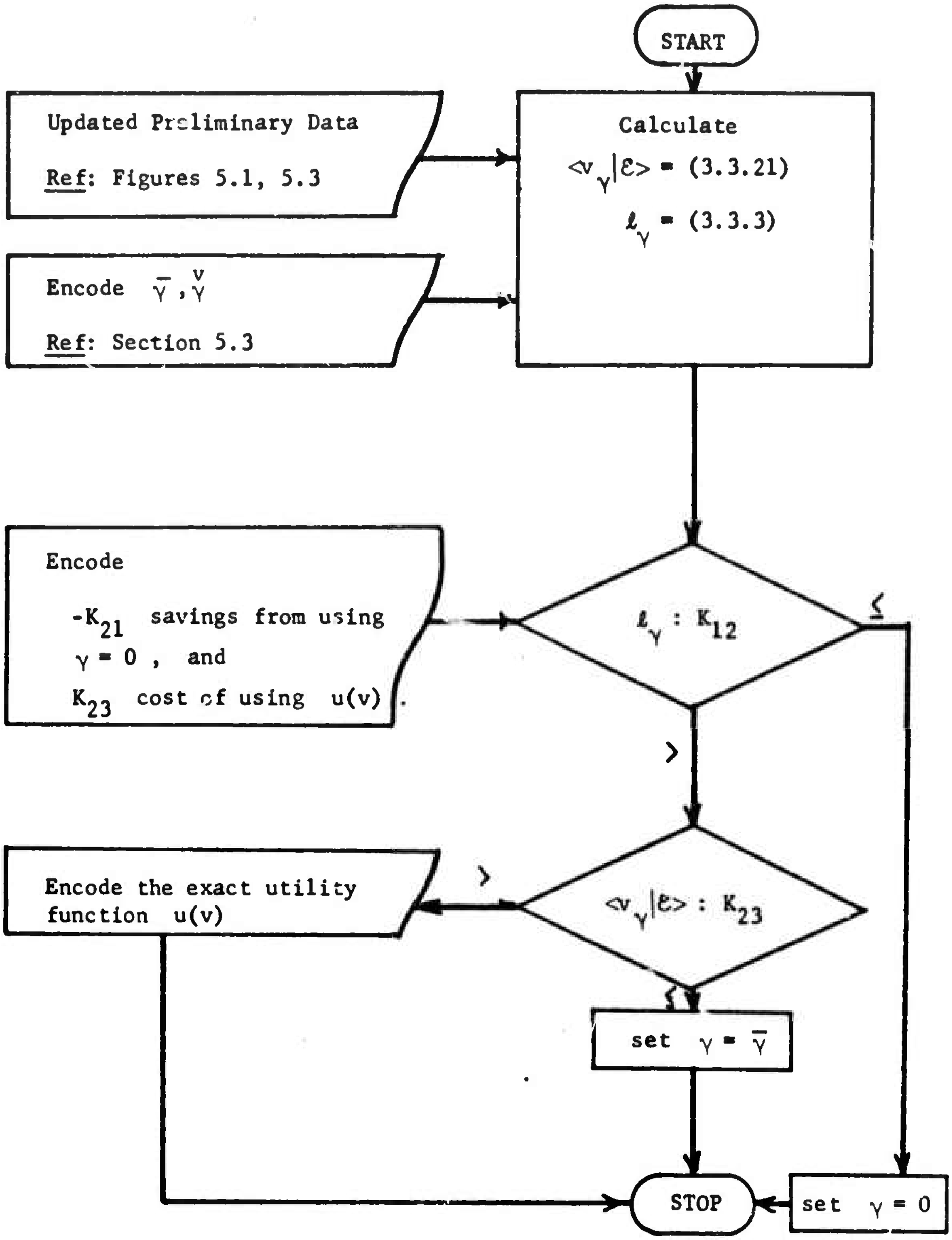


Figure 5.4 Choice of risk preference alternative

that would occur if risk preference were suppressed in the probabilistic phase. K_{23} is the additional cost of encoding the decision maker's complete risk preference function and using it for computations. The logic for comparing costs and gains is shown at the bottom of Fig. 5.4. We assume that

$$K_{23} \geq K_{12} \tag{5.3.1}$$

and

$$\langle v_\gamma | \mathcal{E} \rangle \leq \ell_\gamma \tag{5.3.2}$$

so that level (iii) is not cost effective, if level (ii) is not.

Discussion of Input Data

The practical application of Fig. 5.4 depends on an accurate estimation of the preliminary risk attitude and of the costs. We consider each below.

The preliminary risk attitude can come from several sources. First, we could encode the entire risk preference function from an assistent. Second, the analyst could examine the risk attitudes of similar decision makers. For example, we expect two corporations with the same assets and earnings to have approximately the same risk attitude. Third, and most promising, we can use modeling. Suppose we assume that u(v) is logarithmic :

$$u(v) = \ln(v + a) \tag{5.3.3}$$

Then we can ask the decision maker the single question:

> Suppose you had the opportunity to call the toss of a fair coin. You win a dollars if you call correctly and you lose a/2 dollars if you are incorrect. For what price a are you indifferent between playing and not playing?

The prize a is the same as the parameter a in (5.3.3), so that u(v) is completely specified.

The future risk attitude is described by how much γ varies from γ_0 . We define γ_0 as the coefficient which would make the certain equivalent approximation exact given the correct profit lottery and utility function,

$$\tilde{\langle} v\lfloor\hat{d}(\mathcal{E}),\mathcal{E}\rangle = \langle v\lfloor\hat{d}(\mathcal{E}),\mathcal{E}\rangle - \frac{1}{2}\gamma_0 \,{}^{v}\langle v\lfloor\hat{d}(\mathcal{E}),\mathcal{E}\rangle . \qquad (5.3.4)$$

We identify three ways that γ_0 can vary from γ .

First, the approximate mean is not exact. We might estimate how much the risk aversion coefficient changes with changes in the mean by assuming that the true utility function is logarithmic. The variance of the difference between the accurate and preliminary means of the profit lottery could then be used to impute a variance in γ . The variance of the profit lottery mean is discussed in the next section.

Second, the variance of the profit lottery may be large enough that exponential utility is not a good local approximation. We can check this effect by assuming again that the true utility is logarithmic. Then we can calculate how much we would have to change γ to make the exponential certain equivalent equal to the logarithmic one.

Third, even if the decision maker's risk attitude is adequately estimated by the exponential utility function, our preliminary estimate of the coefficient may be wrong. We can directly encode how much the mean will shift during an encoding session. Just as for state variable encoding, experimental data is helpful in assessing potential mean shifts. Spetzler [8] has made a start in this direction.

Computing the three risk coefficient variations described above should help in assigning $\check{\gamma}$ the prior variance of the posterior risk coefficient.

Costs

The final question is costs. The gains from going from level (ii) to level (i) should not be large. By suppressing risk aversion, the analyst does not have to educate the decision maker or assess γ accurately. Computationally, exponential utility is almost as tractable as linear utility. Exponential utility has the delta property; if a fixed number of dollars δ is added to each prize in a lottery, the certain equivalent of the lottery is increased by δ . Because of the delta property, computation with exponential utility can be decomposed. The profit lottery can be generated without considering risk aversion or utility, and the certain equivalent can be computed later.

By going from level (ii) to level (iii), we find that the costs rise sharply. The education and assessment costs can be large, especially if the decision maker is an organization. Spetzler's [8] study indicates that accurate determination of γ is a time-consuming job. The computational burden is also large because we lose the delta property.

Risk Preference in the Entrepreneur's Problem

The key inputs for the Entrepreneur's Problem are the mean and variance of γ . From Section 5.1 we have

$$\bar{\gamma} = 0.004 \; . \tag{5.3.5}$$

To estimate the variance we compute the asset level which would imply (5.3.5). Combining the definition of the local risk aversion coefficient

with the expression for logarithmic utility we have

$$\gamma(v) = -\frac{u''(v)}{u'(v)} = \frac{1}{v+a} , \tag{5.3.6}$$

where prime denotes differentiation. Evaluating at $\langle v|\hat{d}(\mathcal{E}),\mathcal{E}\rangle = 200.5$ and $\bar{\gamma} = .004$, the asset level a is

$$a = \frac{1}{0.004} - 200.5 = 49.5 . \tag{5.3.7}$$

Figure 5.5 shows the entrepreneur's normalized profit lottery. The expected value if 1.0 standard deviations above zero. The prior assets a are only 0.25 standard deviations. Therefore, if the outcome of the lottery is more than 1.25 standard deviations below the mean, the loss will exceed the entrepreneur's assets, resulting in bankruptcy. Assuming normality, there is an 11% chance of bankruptcy.

Given a finite probability of bankruptcy the logarithmic certain equivalent for the lottery is negative and arbitrarily large. Because the logarithmic and exponential certain equivalent vary widely, the actual risk aversion coefficient may be quite different from the preliminary estimate. A variance that reflects this uncertainty is

$$\check{\gamma} = 0.0004 . \tag{5.3.8}$$

The remaining inputs for Fig. 5.5 are K_{12} and K_{23} . The savings from using level (i) $-K_{12}$ is the cost of encoding the risk aversion coefficient. Like the cost of encoding, this amount is approximately one thousand dollars :

$$-K_{12} = 10^{-3} \tag{5.3.9}$$

We assume that the cost of encoding and using the complete risk

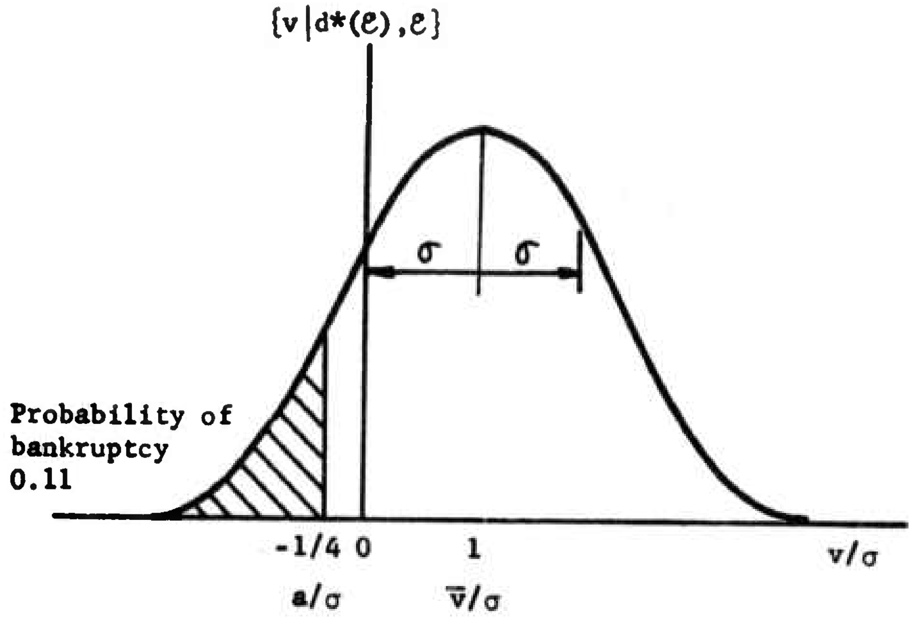


Figure 5.5 **The entrepreneur's profit lottery**

preference function is ten thousand dollars :

$$K_{23} = 10^{-2} \tag{5.3.10}$$

Comparing the costs and benefits as required by the decision boxes of Fig. 5.5, we find that the best decision is clearly to encode and use the complete utility function.

5.4 The Choice of a Computational Alternative

The purpose of computation is to accurately estimate the optimum decision $\hat{d}(\mathcal{E})$. The options considered in this section are to use $\hat{d}(\mathcal{E})^a$ the approximate optimum from Chapter 3 or to find a more accurate estimate through Monte Carlo simulation. The analysis is limited to problems with a single decision variable.

Using the preliminary estimate directly is the best alternative if it is accurate and the cost per evaluation of the deterministic model is high. In this case our preliminary estimate of the optimum decision becomes our actual decision.

Monte Carlo sampling is described in Chapter 4. Random samples are drawn from $\{s|\mathcal{E}\}$. At each decision setting d_j , an approximation to $\langle v|d_j,\mathcal{E}\rangle$ is computed. Then a least squares curve is fit through the means and maximized.

Inability to Estimate Tree Errors

Logically, we would include a decision tree as a computational alternative. Decision trees are generated by discretizing both d and $\{\underline{s}|\mathcal{E}\}$. The optimum decision is computed by rolling back the tree.

Unfortunately, the optimal tree for a quadratic value function has exactly one terminal node. The problem is that errors for trees with two or more branches per state variable depend on partial derivatives

of greater than second order. Since we have suppressed these derivatives by using a quadratic value function, we have lost the ability to design decision trees.

The Evaluation Scheme

The evaluation scheme is shown in Fig. 5.6. The only new input required is the cost per Monte Carlo sample K_e , which normally is dominated by the cost per evaluation of the deterministic model $v(\underline{s},\underline{d})$.

In Chapter 4 we show that the expected loss is a function of the total sample size N , even if d is roughly quantized. If the sample data swamps the prior data, the N^* computed in Fig. 5.6 is the optimal sample size. However, if N^* is small we leave it to the user to decide whether to sample or not.

Discussion

Instead of leaving the final choice to the user, we could encode the prior variance of the posterior mean

$$^v\langle\langle v|N,\mathcal{E}\rangle|\mathcal{E}\rangle \ . \tag{5.4.1}$$

The sampling scheme of Chapter 4 would have to be modified to include prior data. Then the calculation of optimal N^* would include the decision; stop or sample.

However, we feel that encoding the quantity in (5.4.1) is as difficult as directly deciding to sample or not. In most practical analyses we suspect the decision will clearly be to sample. In this case the sample data will swamp the prior data, and our evaluation scheme eliminates unnecessary encoding.

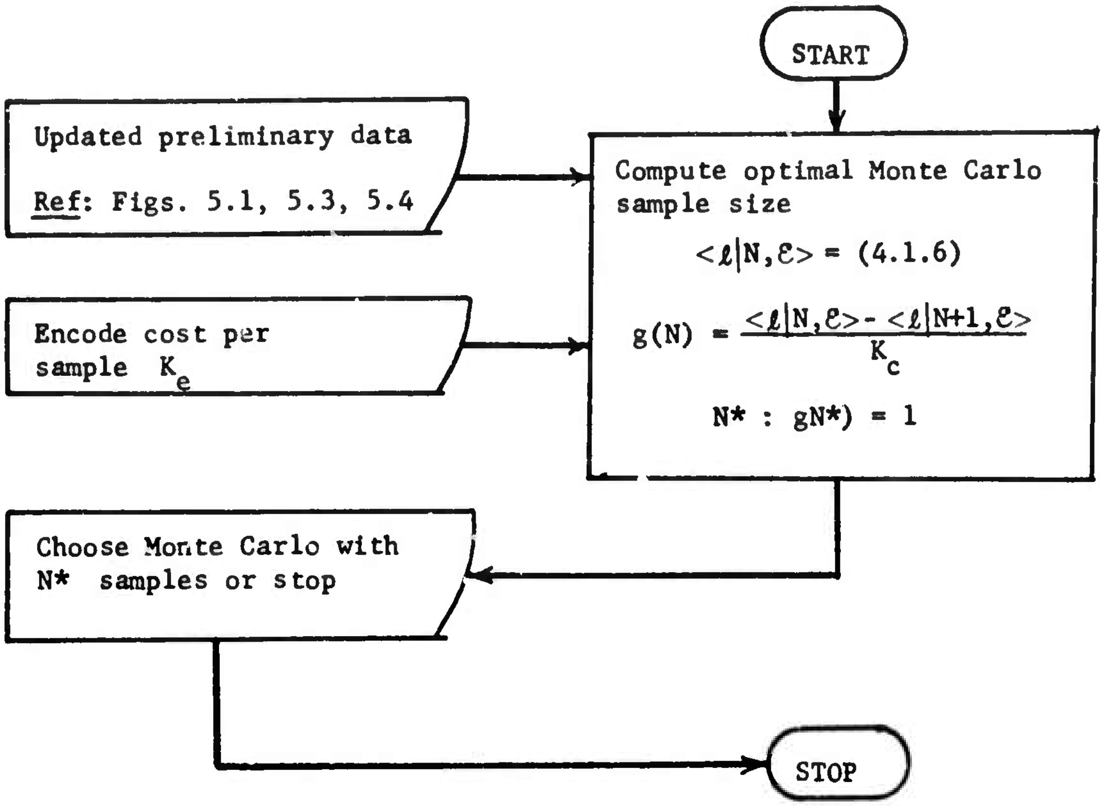


Figure 5.6   The choice of a computational procedure

Computation in the Entrepreneur's Problem

Suppose we decide to compute the Entrepreneur's profit for prices from 15 to 35. If the cost per sample K_c is \$.10, then the optimum sample size N^* is

$$N^* = 820 . \tag{5.4.2}$$

Since we evaluated $v(\underline{s},\underline{d})$ far fewer than 820 times during the preliminary analysis, we expect the Monte Carlo data to swamp our prior knowledge. Therefore, the results of Chapter 4 hold, and the decision is clearly to sample. The cost of the sampling program is \$80, negligible compared to the encoding costs.

Conclusion of the Entrepreneur's Problem

Our framework has given us insight into the Entrepreneur's Problem. Although computation is necessary, it is routine. Encoding the state variable Δq is important. Using the rough estimate of $\{\Delta q|\mathcal{E}\}$ would result in an expected loss of about 5% of the expected value of the lottery. The crucial issue is risk preference. Because the lottery is the entrepreneur's largest asset, we expect risk preference to be the dominant issue. The large value of additional risk assessment, 50% of the expected value of the lottery, indicates that the framework has pinpointed the critical issue.

5.5 Extensions to Budget Constrained Design and to Discrete Decisions

Both budget constrained design and discrete decisions should be straightforward to include within our framework.

For budget constrained design we need to compute an approximate cost benefit ratio for each option. For encoding we compute the expected

value of sample data divided by the cost K_e for each variable. For risk preference we divide the gain from going from level (i) to level (ii) ℓ_γ by the cost $-K_{12}$. Similarly, we compute the ratio of $\langle v_\gamma|\mathcal{E}\rangle$ to K_{23} for the level (ii) to level (iii) transition. For Monte Carlo computation the benefit to cost ratio is computed as a function of the number of samples N in Fig. 5.6.

Budget constrained design would balance the overall effort in each of the three areas. As nearly as the discrete nature of the options would allow, the last variable encoded and the last sample taken would have the same benefit-cost ratio as the risk preference level chosen.

The framework could be modified to treat problems with discrete decision variables by basing the value of information on closed loop sensitivities as in Section 2.5. The risk-preference evaluation would require a closed loop risk sensitivity. The comparison of Monte Carlo and trees is simplified since the decision variable is already discretized. However, tree errors remain difficult to compute.

CHAPTER 6

SUMMARY AND SUGGESTIONS FOR FUTURE WORK

The objective of the thesis was to create a paradigm to evaluate the economic value of analysis. To achieve that goal in Chapter 5, we derived theoretical results in Chapters 2, 3, and 4.

Summary

The theorem of Chapter 2 introduced the key concept that the value of data for a continuous quadratic problem is proportional to the prior covariance of the posterior means of the state variables. We showed that special cases of the theorem are well known. The constant of proportionality in the theorem contained only second order partial derivatives, which could be evaluated from closed loop sensitivities. Using the idea of compensation we derived a methodology which ranks the state variables accurately for a broad class of decision problems. The final conclusion from Chapter 2 was that the loss from deliberate introduction of error had the same form as the value of data with squared means replacing variances.

The theorem of Chapter 3 gave the value of data for the risk-sensitive quadratic problem using the certain equivalent approximation. To operationalize the result we assumed that the state variables were normal and independent. More interesting than the theorem itself were the conditions under which the risk-sensitive value of data reduced to the risk-indifferent value of data. When these conditions held, we found that the risk aversion coefficient could be treated as a state variable for value of data calculations.

Chapter 4 considered the penalty for rough quantization of a decision variable. We found that although fine quantization was superior to rough quantization, the expected loss from rough quantization was negligibly small. The sensitive parameter was sharpness, a measure of the curvature of the expected value over the range of the decision variable.

Chapter 5 applied the results of Chapters 2, 3, and 4. The flowcharts were presented to help the analyst to balance encoding, risk preference, and computation for future decision analyses.

Further Research

The results of this thesis can be regarded as a black box. The inputs are estimates of how something might change during a data generating process, and the output is a dollar valuation of the potential change. The most valuable future research would be a series of boxes which could be attached to the front of our black box. The future boxes would contain experimental and historical data. From data that is easy to encode, the new boxes would generate the input for our black boxes.

For encoding state variables, two boxes would be useful. The input to the first would be elementary data such as means and ranges of state variables. The output would divide the prior variance into the prior expectation of the posterior variance and the prior variance of the posterior mean. Research comparing preliminary estimates to thoroughly encoded probability density functions would be required to generate the box.

The second encoding box would predict the variance of the posterior mean as a function of the length of the encoding interview. Since the

cost should also be proportional to the length of the interview, this box would allow us to compute the value of partial encoding. The required research would include the design of good encoding techniques as well as the recording of mean shifts during encoding sessions.

The box required for risk preference is similar to the second encoding box. It should predict the potential shift of the risk aversion coefficient as a function of the completeness of the encoding.

Computation is probably the most challenging area. A box is needed for probability trees. It should predict the errors as a function of the number of terminal nodes in a tree. The obstacle to be overcome is that the errors in a probability tree are proportional high order partial derivatives of the value function with respect to the state variables. Therefore, the quadratic model of the value function is not appropriate.

The application of the economics of decision analysis depends on careful modeling so that encoding the potential data does not become a burden.

LIST OF REFERENCES

[1] De Groot, Morris. Optimal Statistical Decisions, McGraw-Hill, 1970.

[2] Howard, Ronald A. "The Foundations of Decision Analysis," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3, Sept. 1968.

[3] Howard, Ronald A. "Proximal Decision Analysis," Management Science, Vol. 17, No. 9, May 1971.

[4] Matheson, James E. "The Economic Value of Analysis and Computation," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3, Sept. 1968.

[5] Pratt, John W. "Risk Aversion in the Small and in the Large," Econometrica, Vol. 32, No. 1-2, Jan.-April 1964.

[6] Raiffa, Howard A. Decision Analysis, Introductory Lectures on Choices under Uncertainty, Addison Wesley, 1968.

[7] Raiffa, Howard A., and Schalifer, Robert. Applied Statistical Decision Theory, M.I.T. Press, 2nd Edition, 1968.

[8] Ralston, Anthony. A First Course in Numerical Analysis, McGraw-Hill, 1965.

[9] Spetzler, Carl. "The Development of a Corporate Risk Policy for Capital Investment Decisions," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3, Sept. 1968.